(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 November 2006 (30.11.2006)

PCT

(10) International Publication Number
WO 2006/126858 A2

(51) International Patent Classification:
G10L 19/00 (2006.01)

(21) International Application Number:
PCT/KR2006/002020

(22) International Filing Date: 26 May 2006 (26.05.2006)

(25) Filing Language: Korean

(26) Publication Language: English

(30) Priority Data:

60/684,578	26 May 2005 (26.05.2005)	US
60/758,608	13 January 2006 (13.01.2006)	US
60/787,172	30 March 2006 (30.03.2006)	US
10-2006-0030660	4 April 2006 (04.04.2006)	KR
10-2006-0030661	4 April 2006 (04.04.2006)	KR
10-2006-0030658	4 April 2006 (04.04.2006)	KR
10-2006-0046972	25 May 2006 (25.05.2006)	KR

(71) **Applicant** *(for all designated States except US)*: **LG ELECTRONICS INC.** [KR/KR]; 20, Yoido-dong, Youngdungpo-gu, Seoul 150-721 (KR).

(72) **Inventors; and**
(75) **Inventors/Applicants** *(for US only)*: **OH, Hyen O** [KR/KR]; 306-403, Gangseon Maeul 3-danji APT., Juyeop 1(il)-dong, Ilsan-gu, Goyang-si, Gyeonggi-do 151-057 (KR). **JUNG, Yang Won** [KR/KR]; 202, #287-4, Yeonhui 3(sam)-dong, Seodaemun-gu, Seoul 120-830 (KR). **PANG, Hee Suk** [KR/KR]; 101, #14-10, Yangjae-dong, Seocho-gu, Seoul 137-130 (KR). **KIM, Dong Soo** [KR/KR]; 1502, Woorim Villa, #602-265, Namhyeon-dong, Gwanak-gu, Seoul 151-801 (KR). **LIM, Jae Hyun** [KR/KR]; 609, Parkvill Officetel, #1062-20, Namhyeon-dong, Gwanak-gu, Seoul 151-801 (KR).

(74) **Agents: KIM, Yong In** et al.; 15th Floor Yo Sam Building, 648-23, Yeoksam-dong, Kangnam-gu, Seoul 135-808 (KR).

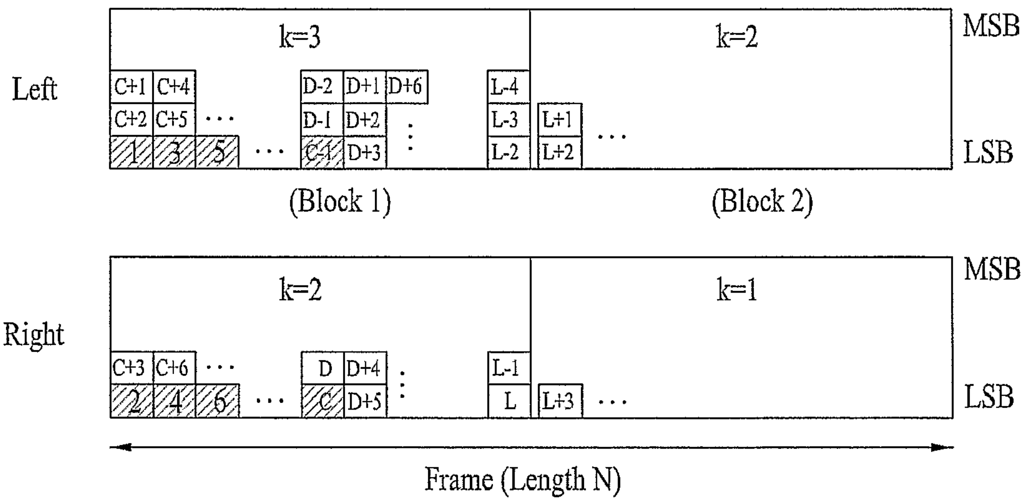
(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

[Continued on next page]

(54) **Title:** METHOD OF ENCODING AND DECODING AN AUDIO SIGNAL



(57) **Abstract:** An apparatus for encoding and decoding an audio signal and method thereof are disclosed, by which compatibility with a player of a general mono or stereo audio signal can be provided in coding an audio signal and by which spatial information for a multi-channel audio signal can be stored or transmitted without a presence of an auxiliary data area. The present invention includes extracting side information embedded in non-recognizable component of audio signal components and decoding the audio signal using the extracted side information.

WO 2006/126858 A2

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Method of Encoding and Decoding an Audio Signal

TECHNICAL FIELD

The present invention relates to a method of encoding and decoding an audio signal.

BACKGROUND ART

Recently, many efforts are made to research and develop various coding schemes and methods for digital audio signals and products associated with the various coding schemes and methods are manufactured.

And, coding schemes for changing a mono or stereo audio signal into multi-channel audio signal using spatial information of the multi-channel audio signal have been developed.

However, in case of storing an audio signal in some recording media, an auxiliary data area for storing spatial information does not exist. So, in this case, only a mono or stereo audio signal is reproduced because the mono or stereo audio signal is stored or transmitted. Hence, a sound quality is monotonous.

Moreover, in case of storing or transmitting spatial information separately, there exists a problem of compatibility with a player of a general mono or stereo audio signal.

DISCLOSURE OF THE INVENTION

Accordingly, the present invention is directed to an apparatus for encoding and decoding an audio signal and method thereof that substantially obviate one or more of the problems due to limitations and disadvantages of the related art.

An object of the present invention is to provide an apparatus for encoding and decoding an audio signal and method thereof, by which compatibility with a player of a general mono or stereo audio signal can be provided in coding an audio signal.

Another object of the present invention is to provide an apparatus for encoding and decoding an audio signal and method thereof, by which spatial information for a multi-channel audio signal can be stored or transmitted without a presence of an auxiliary data area.

Additional features and advantages of the present invention will be set forth in the description which follows, and in part will be apparent from the description, or may be learned by practice of the invention. The objectives and other advantages of the present invention will be realized and attained by the structure particularly pointed out in the written description and claims thereof as well as the appended drawings.

To achieve these and other advantages and in accordance with the purpose of the present invention, a method of decoding an audio signal according to the present invention includes a step (a) of extracting side information embedded in the audio signal by being dispersed on at least one channel of the audio signal and a step (b) of decoding the audio signal using the side information.

To further achieve these and other advantages and in accordance with the purpose of the present invention, a method of encoding an audio signal according to the present invention includes a step (a) of generating side information necessary for decoding an audio signal and a step (b) of embedding the side information in the audio signal having at least one channel by dispersing the side information.

To further achieve these and other advantages and in accordance with the purpose of the present invention, a data structure according to the present invention includes an audio signal and side information necessary for decoding the audio signal embedded in non-recognizable components of the audio signal having at least one channel by being dispersed.

To further achieve these and other advantages and in accordance with the purpose of the present invention, an apparatus for encoding an audio signal according to the present invention includes a side information generating unit for generating side information necessary for decoding an audio signal and an embedding unit for embedding the side information in the audio signal having at least one channel by dispersing the side information.

To further achieve these and other advantages and in accordance with the purpose of the present invention, an apparatus for decoding an audio signal according to the present invention includes an embedded signal decoding unit for extracting side information embedded in the audio signal having at least one channel by being dispersed and a multi-channel generating unit for decoding the audio signal using the additional information.

It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory and are intended to provide further explanation of the invention as claimed.

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are included to provide a further understanding of the invention and are incorporated in and constitute a part of this specification, illustrate embodiments of the invention and together with the description serve to explain the principles of the invention.

In the drawings:

FIG. 1 is a diagram for explaining a method that a human recognizes spatial information for an audio signal according to the present invention;

FIG. 2 is a block diagram of a spatial encoder according to the present invention;

FIG. 3 is a detailed block diagram of an embedding unit configuring the spatial encoder shown in FIG. 2 according to the present invention;

FIG. 4 is a diagram of a first method of rearranging a spatial information bitstream according to the present invention;

FIG. 5 is a diagram of a second method of rearranging a spatial information bitstream according to the present invention;

FIG. 6A is a diagram of a reshaped spatial information bitstream according to the present invention;

FIG. 6B is a detailed diagram of a configuration of the spatial information bitstream shown in FIG. 6A;

FIG. 7 is a block diagram of a spatial decoder according to the present invention;

FIG. 8 is a detailed block diagram of an embedded signal decoder included in the spatial decoder according to the present invention;

FIG. 9 is a diagram for explaining a case that a general PCM decoder reproduces an audio signal according to the present invention;

FIG. 10 is a flowchart of an encoding method for embedding spatial information in a downmix signal according to the present invention;

FIG. 11 is a flowchart of a method of decoding spatial information embedded in a downmix signal according to the present invention;

FIG. 12 is a diagram for a frame size of a spatial information bitstream embedded in a downmix signal according to the present invention;

FIG. 13 is a diagram of a spatial information bitstream embedded by a fixed size in a downmix signal according to the present invention;

FIG. 14A is a diagram for explaining a first method for solving a time align problem of a spatial information bitstream embedded by a fixed size;

FIG. 14B is a diagram for explaining a second method for solving a time align problem of a spatial information bitstream embedded by a fixed size;

FIG. 15 is a diagram of a method of attaching a spatial information bitstream to a downmix signal according to the present invention;

FIG. 16 is a flowchart of a method of encoding a spatial information bitstream embedded by various sizes in a downmix signal according to the present invention;

FIG. 17 is a flowchart of a method of encoding a spatial information bitstream embedded by a fixed size in a downmix signal according to the present invention;

FIG. 18 is a diagram of a first method of embedding a spatial information bitstream in an audio signal downmixed on at least one channel according to the present invention;

FIG. 19 is a diagram of a second method of embedding a spatial information bitstream in an audio signal downmixed on at least one channels according to the present invention;

FIG. 20 is a diagram of a third method of embedding a spatial information bitstream in an audio signal downmixed on at least one channel according to the present invention;

FIG. 21 is a diagram of a fourth method of embedding a spatial information bitstream in an audio signal downmixed on at least one channel according to the present invention;

FIG. 22 is a diagram of a fifth method of embedding a spatial information bitstream in an audio signal downmixed on at least one channel according to the present invention;

FIG. 23 is a diagram of a sixth method of embedding a spatial information bitstream in an audio signal downmixed on at least one channel according to the present invention;

FIG. 24 is a diagram of a seventh method of embedding a spatial information bitstream in an audio signal downmixed on at least one channel according to the present invention;

FIG. 25 is a flowchart of a method of encoding a spatial information bitstream to be embedded in an audio signal downmixed on at least one channel according to the present invention; and

FIG. 26 is a flowchart of a method of decoding a spatial information bitstream embedded in an audio signal downmixed on at least one channel according to the present invention.

BEST MODE FOR CARRYING OUT THE INVENTION

Reference will now be made in detail to the preferred embodiments of the present invention, examples of which are illustrated in the accompanying drawings.

First of all, the present invention relates to an apparatus for embedding side information necessary for decoding an audio signal in the audio signal and method thereof. For the convenience of explanation, the audio signal and side information are represented as a downmix signal and spatial information in the following description, respectively, which does not put limitation on the present invention. In this case, the audio signal includes a PCM signal.

FIG. 1 is a diagram for explaining a method that a human recognizes spatial information for an audio signal according to the present invention

Referring to FIG. 1, based on a fact that a human is able to recognize an audio signal 3-dimensionally, a coding scheme for a multi-channel audio signal uses a fact that the audio signal can be represented as 3-dimensional spatial information via a plurality of parameter sets.

Spatial parameters for representing spatial information of a multi-channel audio signal include CLD (channel level differences), ICC (inter-channel coherences), CTD (channel time difference), etc. The CLD means an energy difference between two channels, the ICC means a correlation between two channels, and the CTD means a time difference between two channels.

How a human recognizes an audio signal spatially and how a concept of the spatial parameter is generated are explained with reference to FIG. 1.

A direct sound wave 103 arrives at a left ear of a human from a remote sound source 101, while another direct sound wave 102 is diffracted around a head to reach a right ear 106 of the human.

The two sound waves 102 and 103 differ from each other in arriving time and energy level. And, the CTD and CLD parameters are generated by using theses differences.

If reflected sound waves 104 and 105 arrive at both of the ears, respectively or if the sound source is dispersed, sound waves having no correlation in-between will arrive at both of the ears, respectively to generate the ICC parameter.

Using the generated spatial parameters according to the above-explained principle, it is able to transmit a multi-channel audio signal as a mono or stereo signal and to output the signal into a multi-channel signal.

The present invention provides a method of embedding the spatial information, i.e., the spatial parameters in the mono or stereo audio signal, transmitting the embedded signal, and reproducing the transmitted signal into a multi-channel audio signal. The present invention is not limited to the multi-channel audio signal. In the following description of the present invention, the multi-channel audio signal is explained for the convenience of explanation.

FIG. 2 is a block diagram of an encoding apparatus according to the present invention.

Referring to FIG. 2, the encoding apparatus according to the present invention receives a multi-channel audio signal 201. In this case, 'n' indicates the number of input channels.

The multi-channel audio signal 201 is converted to a downmix signal (Lo and Ro) 205 by an audio signal generating unit 203. The downmix signal includes a mono or stereo audio signal and can be a multi-channel audio signal. In the present invention, the stereo audio signal will be taken as an example in the following description. Yet, the present invention is not limited to the stereo audio signal.

Spatial information of the multi-channel audio signal, i.e., a spatial parameter is generated from the multi-channel audio signal 201 by a side information generating unit 204. In the present invention, the spatial information indicates information for an audio signal channel used in transmitting the downmixed signal 205 generated by downmixing a multi-channel (e.g., left, right, center, left surround, right surround, etc.) signal and upmixing the transmitted downmix signal into the multi-channel audio signal again. Optionally, the downmix signal 205 can be generated using a downmix signal directly provided from outside, e.g., an artistic downmix signal 202.

The spatial information generated in the side information generating unit 204 is encoded into a spatial information bitstream for transmission and storage by an side information encoding unit 206.

The spatial information bitstream is appropriately reshaped to be directly inserted in an audio signal, i.e., the downmix signal 205 to be transmitted by an embedding unit 207. In doing so, 'digital audio embedded method' is usable.

For instance, in case that the downmix signal 205 is a raw PCM audio signal to be stored in a storage medium (e.g., stereo compact disc) difficult to store the spatial information therein or to be transmitted by SPDIF (Sony/Philips Digital Interface), an auxiliary data field for storing the spatial information does not exist unlike the case of compression encoding by AAC or the like.

In this case, if the 'digital audio embedded method' is used, the spatial information can be embedded in the raw PCM audio signal without sound quality distortion. And, the audio signal having the spatial information embedded therein is not discriminated from the raw signal in aspect of a general decoder. Namely, an output signal Lo'/Ro' 208 having the spatial information embedded therein can be regarded as a same signal of the input signal Lo/Ro 205 in aspect of a general PCM decoder.

As the 'digital audio embedded method', there is a 'bit replacement coding method', an 'echo hiding method', a 'spread-spectrum based method' or the like.

The bit replacement coding method is a method of inserting specific information by modifying lower bits of a quantized audio sample. In an audio signal, modification of lower bits almost has no influence on a quality of the audio signal.

The echo hiding method is a method of inserting an echo small enough not to be heard by human ears in an audio signal.

And, the spread-spectrum based method is a method of transforming an audio signal into a frequency domain via discrete cosine transform, discrete Fourier transform or the like, performing spread spectrum on specific binary information into PN (pseudo noise) sequence, and adding it to the audio signal transformed into the frequency domain.

In the present invention, the bit replacement coding method will be mainly explained in the following description. Yet, the present invention is not limited to the bit replacement coding method.

FIG. 3 is a detailed block diagram of an embedding unit configuring the spatial encoder shown in FIG. 2 according to the present invention.

Referring to FIG. 3, in embedding spatial information in non-perceptive components of downmix signal components by the bit replacement coding method, an insertion bit length (hereinafter named 'K-value') for embedding the spatial information can use K-bit (K>0) according to a pre-decided method instead of using a lower 1-bit only. The K-bit can use lower bits of the downmix signal but is not limited to the lower bits only. In this case, the pre-decided method is a method of finding a masking threshold according to a psychoacoustic model and allocating a suitable bit according to the masking threshold for example.

A downmix signal Lo/Ro 301, as shown in the drawing, is transferred to an audio signal encoding unit 306 via a buffer 303 within the embedding unit.

A masking threshold computing unit 304 segments an inputted audio signal into predetermined sections (e.g., blocks) and then finds a masking threshold for the corresponding section.

The masking threshold computing unit 304 finds an insertion bit length (i.e., K value) of the downmix signal enabling a modification without occurrence of aural distortion according to the masking threshold. Namely, a bit number usable in embedding the spatial information in the downmix signal is allocated per block.

In the description of the present invention, a block means a data unit inserted using one insertion bit length (i.e., K value) existing within a frame.

At least one or more blocks can exist within one frame. If a frame length is fixed, a block length may decrease according to the increment of the number of blocks.

Once the K value is determined, it is able to include the K value in a spatial information bitstream. Namely, a bitstream reshaping unit 305 is able to reshape the spatial information bitstream in a manner of enabling the spatial information bitstream to include the K value therein. In this case, a sync word, an error detection code, an error correction code and the like can be included in the spatial information bitstream.

The reshaped spatial information bitstream can be rearranged into an embeddable form. The rearranged spatial information bitstream is embedded in the downmix signal by an audio signal encoding unit 306 and is then outputted as an audio signal Lo'/Ro' 307 having the spatial information bitstream embedded therein. In this case, the spatial information bitstream can be embedded in K-bits of the downmix signal. The K value can have one fixed value in a block. In any cases, the K value is inserted in the spatial information bitstream in the reshaping or rearranging process of the spatial information bitstream and is then transferred to a decoding apparatus. And, the decoding apparatus is able to extract the spatial information bitstream using the K value.

As mentioned in the foregoing description, the spatial information bitstream goes through a process of being embedded in the downmix signal per block. The process is performed by one of various methods.

A first method is carried out in a manner of substituting lower K bits of the downmix signal with zeros simply and adding the rearranged spatial information bitstream data. For instance, if a K value is 3, if sample data of a downmix signal is 11101101 and if spatial information bitstream data to embed is 111, lower 3 bits of '11101101' are substituted with zeros to provide 11101000. And, the spatial information bitstream data '111' is added to '11101000' to provide '11101111'.

A second method is carried out using a dithering method. First of all, the rearranged spatial information bitstream data is subtracted from an insertion area of the downmix signal. The downmix signal is then re-quantized based on the K value. And, the rearranged spatial information bitstream data is added to the re-quantized downmix signal. For instance, if a K value is 3, if sample data of a downmix signal is 11101101 and if spatial information bitstream data to embed is 111, '111' is subtracted from the '11101101' to provide 11100110. Lower 3 bits are then re-quantized to provide '11101000' (by rounding off). And, the '111' is added to '11101000' to provide '11101111'.

Since a spatial information bitstream embedded in the downmix signal is a random bitstream, it may not have a white-noise characteristic. Since addition of a white-noise type signal to a downmix signal is advantageous in sound quality characteristics, the spatial information bitstream goes through a whitening process to be added to the downmix signal. And, the whitening process is applicable to spatial information bitstreams except a sync word.

In the present invention, 'whitening' means a process of making a random signal having an equal or almost similar sound quantity of an audio signal in all areas of a frequency domain.

Besides, in embedding a spatial information bitstream in a downmix signal, aural distortion can be minimized by applying a noise shaping method to the spatial information bitstream.

In the present invention, 'noise shaping method' means a process of modifying a noise characteristic to enable energy of a quantized noise generated from quantization to move to a high frequency band over an audible frequency band or a process of generating a time-varying filer corresponding to a masking threshold obtained from a corresponding audio signal and modifying a characteristic of a noise generated from quantization by the generated filter.

FIG. 4 is a diagram of a first method of rearranging a spatial information bitstream according to the present invention.

Referring to FIG. 4, as mentioned in the foregoing description, the spatial information bitstream can be rearranged into an embeddable form using the K value. In this case, the spatial information bitstream can be embedded in the downmix signal by being rearranged in various ways. And, FIG. 4 shows a method of embedding the spatial information in a sample plane order.

The first method is a method of rearranging the spatial information bitstream in a manner of dispersing the spatial information bitstream for a corresponding block by K-bit unit and embedding the dispersed spatial information bitstream sequentially.

If a K value is 4 and if one block 405 is constructed with N samples 403, the spatial information bitstream 401 can be rearranged to be embedded in lower 4 bits of each sample sequentially.

As mentioned in the foregoing description, the present invention is not limited to a case of embedding a spatial information bitstream in lower 4 bits of each sample.

Besides, in lower K bits of each sample, the spatial information bitstream, as shown in the drawing, can be embedded in MSB (most significant bit) first or LSB (least significant bit) first.

In FIG. 4, an arrow 404 indicates an embedding direction and a numeral within parentheses indicates a data rearrangement sequence.

A bit plane indicates a specific bit layer constructed with a plurality of bits.

In case that a bit number of a spatial information bitstream to be embedded is smaller than an embeddable bit number in an insertion area in which the spatial information bitstream will be embedded, remaining bits are padded up with zeros 406, a random signal is inserted in the remaining bits, or the remaining bits can be replaced by an original downmix signal.

For instance, if a number (N) of samples configuring a block is 100 and if a K value is 4, a bit number (W) embeddable in the block is W = N*K = 100*4 = 400.

If a bit number (V) of a spatial information bitstream to be embedded is 390 bits (i.e., V<W), remaining 10 bits are padded up with zeros, a random signal is inserted in the remaining 10 bits, or the remlinging 10 bits are replaced by an original downmix signal, the remaining 10 bits are filled up with a tail sequence indicating a data end, or the remaining 10 bits can be filled up with combinations of them. The tail sequence means a bit sequence indicating an end of a spatial information bitstream in a corresponding block. Although Fig. 4 shows that the remaining bits are padded per block, the present invention includes a case that the remaining bits are padded up per insertion frame in the above manner.

FIG. 5 is a diagram of a second method of rearranging a spatial information bitstream according to the present invention.

Referring to FIG. 5, the second method is carried out in a manner of rearranging a spatial information bitstream

501 in a bit plane 502 order. In this case, the spatial information bitstream can be sequentially embedded from a lower bit of a downmix signal per block, which does not put limitation of the present invention.

For instance, if a number (N) of samples configuring a block is 100 and if a K value is 4, 100 least significant bits configuring the bit plane-0 502 are preferentially padded and 100 bits configuring the bit plane-1 502 can be padded.

In FIG. 5, an arrow 505 indicates an embedding direction and a numeral within parentheses indicates a data rearrangement order.

The second method can be specifically advantageous in extracting a sync word at a random position. In searching for the sync word of the inserted spatial information bitstream from the rearranged and encoded signal, only LSB can be extracted to search for the sync word.

And, it can be expected that the second method uses minimum LSB only according to a bit number (V) of a spatial information bitstream to be embedded. In this case, if a bit number (V) of a spatial information bitstream to be embedded is smaller than an embeddable bit number (W) in an insertion area in which the spatial information bitstream will be embedded, remaining bits are padded up with zeros

506, a random signal is inserted in the remaining bits, the remaining bits are replaced by an original downmix signal, the remaining bits are padded with an end bit sequence indicating an end of data, or the remaining bits can be padded with combinations of them. In particular, the method of using the downmix signal is advantageous. Although, FIG. 5 shows an example of padding the remaining bits per block, the present invention includes a case of padding the remaining bits per insertion frame in the above-explained manner.

FIG. 6A shows a bitstream structure to embed a spatial information bitstream in a downmix signal according to the present invention.

Referring to FIG. 6A, a spatial information bitstream 607 can be rearranged by the bitstream reshaping unit 305 to include a sync word 603 and a K value 604 for the spatial information bitstream.

And, at least one error detection code or error correction code 606 or 608 (hereinafter, the error detection code will be described) can be included in the reshaped spatial information bitstream in the reshaping process. The error detection code is capable of deciding whether the spatial information bitstream 607 is distorted in a process of transmission or storage

The error detection code includes CRC (cyclic redundancy check). The error detection code can be included by being divided into two steps. An error detection code-1 for a header 601 having K values and an error detection code-2 for a frame data 602 of the spatial information bitstream can be separately included in the spatial information bitstream. Besides, the rest information 605 can be separately included in the spatial information bitstream. And, information for a rearrangement method of the spatial information bitstream and the like can be included in the rest information 605.

FIG. 6B is a detailed diagram of a configuration of the spatial information bitstream shown in FIG. 6A. FIG. 6B shows an embodiment that one frame of a spatial information bitstream 601 includes two blocks, to which the present invention is not limited.

Referring to FIG. 6B, a spatial information bitstream shown in FIG. 6B includes a sync word 612, K values (K1, K2, K3, K4) 613 to 616, a rest information 617 and error detection codes 618 and 623.

The spatial information bitstream 610 includes a pair of blocks. In case of a stereo signal, a block-1 can be consist of blocks 619 and 620 for left and right channels, respectively. And, a block-2 can be consist of blocks 621 and 62 for left and right channels, respectively.

Although a stereo signal is shown in FIG. 6B, the present invention is not limited to the stereo signal.

Insertion bit lengths (K values) for the blocks are included in a header part.

The K1 613 indicates the insertion bit length for the left channel of the block-1. The K2 614 indicates the insertion bit length of the right channel of the block-1. The K3 615 indicates the insertion bit length for the left channel of the block-2. And, the K4 616 indicates the insertion bit size for the right channel of the block-2.

And, the error detection code can be included by being divided into two steps. For instance, an error detection code-1 618 for a header 609 including the K values therein and an error detection code-2 for a frame data 611 of the spatial information bitstream can be separately included.

FIG. 7 is a block diagram of a decoding apparatus according to the present invention.

Referring to FIG. 7, a decoding apparatus according to the present invention receives an audio signal Lo'/Ro' 701 in which a spatial information bitstream is embedded.

The audio signal having the spatial information bitstream embedded therein may be one of mono, stereo and multi-channel signals. For the convenience of explanation, the stereo signal is taken as an example of the present invention, which does not put limitation on the present invention.

An embedded signal decoding unit 702 is able to extract the spatial information bitstream from the audio signal 701.

The spatial information bitstream extracted by the embedded signal decoding unit 702 is an encoded spatial information bitstream. And, the encoded spatial information bitstream can be an input signal to a spatial information decoding unit 703.

The spatial information decoding unit 703 decodes the encoded spatial information bitstream and then outputs the decoded spatial information bitstream to a multi-channel generating unit 704.

The multi-channel generating unit 704 receives the downmix signal 701 and spatial information obtained from the decoding as inputs and then outputs the received inputs as a multi-channel audio signal 705.

FIG. 8 is a detailed block diagram of the embedded signal decoding unit 702 configuring the decoding apparatus according to the present invention.

Referring to FIG. 8, an audio signal Lo’/Ro’, in which spatial information is embedded, is inputted to the embedded signal decoding unit 702. And, a sync word searching unit 802 detects a sync word from the audio signal 801. In this case, the sync word can be detected from one channel of the audio signal.

After the sync word has been detected, a header decoding unit 803 decodes a header area. In this case, information of a predetermined length is extracted from the header area and a data reverse-modifying unit 804 is able to apply an reverse-whitening scheme to header area information excluding the sync word from the extracted information.

Subsequently, length information of the header area and the like can be obtained from the header area information having the reverse-whitening scheme applied thereto.

And, the data reverse-modifying unit 804 is able to apply the reverse-whitening scheme to the rest of the spatial information bitstream. Information such as a K value and the like can be obtained through the header decoding. An original spatial information bitstream can be obtained by arranging the rearranged spatial information bitstream again using the information such as K value and the like. Moreover, sync position information for arranging frames of a downmix signal and the spatial information bitstream, i.e., a frame arrangement information 806 can be obtained.

FIG. 9 is a diagram for explaining a case that a general PCM decoding apparatus reproduces an audio signal according to the present invention.

Referring to FIG. 9, an audio signal Lo'/Ro', in which a spatial information bitstream is embedded, is applied as an input of a general PCM decoding apparatus.

The general PCM decoding apparatus recognizes the audio signal Lo'/Ro', in which a spatial information bitstream is embedded, as a normal stereo audio signal to reproduce a sound. And, the reproduced sound is not discriminated from an audio signal 902 prior to the embedment of spatial information in aspect of quality of sound.

Hence, the audio signal, in which the spatial information is embedded, according to the present invention has compatibility for normal reproduction of stereo signals in the general PCM decoding apparatus and an advantage in providing a multi-channel audio signal in a decoding apparatus capable of multi-channel decoding.

FIG. 10 is a flowchart of an encoding method for embedding spatial information in a downmix signal according to the present invention.

Referring to FIG. 10, an audio signal is downmixed from a multi-channel signal (1001, 1002). In this case, the downmix signal can be one of mono, stereo and multi-channel signals.

Subsequently, spatial information is extracted from the multi-channel signal (1003). And, a spatial information bitstream is generated using the spatial information (1004).

The spatial information bitstream is embedded in the downmix signal (1005).

And, a whole bitstream including the downmix signal having the spatial information bitstream embedded therein is transferred to a decoding apparatus (1006).

In particular, the present invention finds an insertion bit length (i.e., K value) of an insertion area, in which the spatial information bitstream will be embedded, using the downmix signal and may embed the spatial information bitstream in the insertion area.

FIG. 11 is a flowchart of a method of decoding spatial information embedded in a downmix signal according to the present invention.

Referring to FIG. 11, a decoding apparatus receives a whole bitstream including a downmix signal having a spatial information bitstream embedded therein (1101) and extract the downmix signal from the bitstream (1102).

The decoding apparatus extractes and decodes the spatial information bitstream from the whole bitstream (1103).

The decoding apparatus extracts spatial information through the decoding (1104) and then decodes the downmix signal using the extracted spatial information (1105). In this case, the downmix signal can be decoded into two channels or multi-channels.

In particular, the present invention can extract information for an embedding method of the spatial information bitstream and information of a K value and can decode the spatial information bitstream using the extracted embedding method and the extracted K value.

FIG. 12 is a diagram for a frame length of a spatial information bitstream embedded in a downmix signal according to the present invention.

Referring to FIG. 12, a 'frame' means a unit having one header and enabling an independent decoding of a predetermined length. In the description of the present invention, a 'frame' means an 'insertion frame' that is going to come next. In the present invention, an 'insertion frame' means a unit of embedding a spatial information bitstream in a downmix signal.

And, a length of the insertion frame can be defined per frame or can use a predetermined length.

For instance, the insertion frame length is made to become a same length of a frame length (s) (hereinafter called 'decoding frame length) of a spatial information bitstream corresponding to a unit of decoding and applying spatial information (cf. (a) of FIG. 12), to become a multiplication of 'S' (cf. (b) of FIG. 12), or to enable 'S' to become a multiplication of 'N' (cf. (c) of FIG. 12).

In case of N=S, as shown in (a) of FIG. 12, the decoding frame length (S, 1201) coincides with the insertion frame length (N, 1202) to facilitate a decoding process.

In case of $N>S$, as shown in (b) of FIG. 12, it is able to reduce a number of bits attached due to a header, an error detection code (e.g., CRC) or the like in a manner of transferring one insertion frame (N, 1204) by attaching a plurality of decoding frames (1203) together.

In case of $N<S$, as shown in (c) of FIG. 12, it is able to configure one decoding frame (S, 1205) by attaching several insertion frames (N, 1206) together.

In the insertion frame header, information for an insertion bit length for embedding spatial information therein, information for the insertion frame length (N), information for a number of subframes included in the insertion frame or the like can be inserted.

FIG. 13 is a diagram of a spatial information bitstream embedded in a downmix signal by an insertion frame unit according to the present invention.

First of all, in each of the cases shown in (a), (b) and (c) of FIG. 12, the insertion frame and the decoding frame are configured to be a multiplication from each other.

Referring to FIG. 13, for transferring, it is able to configure a bitstream of a fixed length, e.g., an packet in such a format as a transport stream (TS) 1303.

In particular, a spatial information bitstream 1301 can be bound by a packet unit of a predetermined length regardless of a decoding frame length of the spatial information bitstream. The packet in which information such as a TS header 1302 and like is inserted can be transferred to a decoding apparatus. A length of the insertion frame can be defined per frame or can use a predetermined length instead of being defined within a frame.

This method is necessary to vary a data rate of a spatial information bitstream by considering that a masking threshold differs per block according to characteristics of a downmix signal and a maximum bit number (K_max) that can be allocated without sound quality distortion of the downmix signal is different.

For instance, in case that the K_max is insufficient to entirely represent a spatial information bitstream needed by a corresponding block, data is transferred up to K_max and the rest is transferred later via another block.

In the K_max is sufficient, a spatial information bitstream for a next block can be loaded in advance.

In this case, each TS packet has an independent header. And, a sync word, TS packet length information, information for a number of subframes included in TS packet, information for insertion bit length allocated within a packet or the like can be included in the header.

FIG. 14A is a diagram for explaining a first method for solving a time align problem of a spatial information bitstream embedded by an insertion frame unit.

Referring to FIG. 14A, a length of an insertion frame is defined per frame or can use a predetermined length.

An embedding method by an insertion frame unit may cause a problem of a time alignment between an insertion frame start position of an embedded spatial information bitstream and a downmix signal frame. So, a solution for the time alignment problem is needed.

In the first method shown in FIG. 14A, a header 1402 (hereinafter called 'decoding frame header') for a decoding frame 1403 of spatial information is separately placed.

Discriminating information indicating whether there exists position information of an audio signal to which the spatial information will be applied can be included within the decoding frame header 1402.

For instance, in case of a TS packet 1404 and 1405, a discriminating information 1408 (e.g., flag) indicating whether there exists the decoding frame header 1402 can be included in the TS packet header 1404.

If the discriminating information 1408 is 1, i.e., if the decoding frame header 1402 exists, the discriminating information indicating whether position information of a downmix signal to which the spatial information bitstream will be applied can be extracted from the decoding frame header.

Subsequently, position information 1409 (e.g., delay information) for the downmix signal to which the spatial information bitstream will be applied, can be extracted from the decoding frame header 1402 according to the extracted discriminating information.

If the discriminating information 1411 is 0, the position information may not be included within the header of the TS packet.

In general, the spatial information bitstream 1403 preferably comes ahead of the corresponding downmix signal 1401. So, the position information 1409 could be a sample value for a delay.

Meanwhile, in order to prevent a problem that a quantity of information necessary for representing the sample value excessively increases due to the delay that is excessively large, a sample group unit (e.g., granule unit) for representation of a group of samples or the like is defined. So, the position information can be represented by the sample group unit.

As mentioned in the foregoing description, a TS sync word 1406, an insertion bit length 1407, the discriminating information indicating whether there exists the decoding frame header and the rest information 140 can be included within the TS header.

FIG. 14B is a diagram for explaining a second method for solving a time align problem of a spatial information bitstream embedded by an insertion frame having a length defined per frame.

Referring to FIG. 14B, in case of a TS packet for example, the second method is carried out in a manner of matching a start point 1413 of a decoding frame, a start point of the TS packet and a start point of a corresponding downmix signal 1412.

For the matched part, discriminating information 1420 or 1422 (e.g., flag) indicating that the three kinds of the start points are aligned can be included within a header 1415 of the TS packet.

FIG. 14B shows that the three kinds of start points are matched at an n^{th} frame 1412 of a downmix signal. In this case, the discriminating information 1422 can have a value of 1.

If the three kinds of start points are not matched, the discriminating information 1420 can have a value of 0.

To match the three kinds of the start points together, a specific portion 1417 next to a previous TS packet is padded up with zeros, has a random signal inserted therein, is replaced by an originally downmixed audio signal or is padded up with combinations of them.

As mentioned in the foregoing description, a TS sync word 1418, an insertion bit length 1419 and the rest information 1421 can be included within the TS packet header 1415.

FIG. 15 is a diagram of a method of attaching a spatial information bitstream to a downmix signal according to the present invention.

Referring to FIG. 15, a length of a frame (hereinafter called 'attaching frame') to which a spatial information bitstream is attached can be a length unit defined per frame or a predetermined length unit not defined per frame.

For instance, an insertion frame length, as shown in the drawing, can be obtained by multiplying or dividing a decoding frame length 1504 of spatial information with N, wherein N is a positive integer or the insertion frame length can have a fixed length unit.

If the decoding frame length 1504 is different from the insertion frame length, it is able to generate the insertion frame having the same length as the decoding frame length 1504, for example, without segmenting the spatial information bitstream instead of cutting the spatial information bitstream randomly to be fitted into the insertion frame.

In this case, the spatial information bitstream can be configured to be embedded in a downmix signal or can be configured to be attached to the downmix signal instead of being embedded in the downmix signal.

In such a signal (hereinafter called a 'first audio signal') as a PCM signal, which is converted to a digital signal from an analog signal, the spatial information bitstream can be configured to be embedded in the first audio signal.

In such a more compressed digital signal (hereinafter called a 'second audio signal') as an MP3 signal, the spatial information bitstream can be configured to be attached to the second audio signal.

In case of using the second audio signal, for example, the downmix signal can be represented as a bitstream in a compressed format. So, a downmix signal bitstream 1502, as shown in the drawing, exists in a compressed format and the spatial information of the decoding frame length 1504 can be attached to the downmix signal bitstream 1502.

Hence, the spatial information bitstream can be transferred at a burst.

A header 1503 can exist in the decoding frame. And, position information of a downmix signal to which spatial information is applied can be included in the header 1503.

Meanwhile, the present invention includes a case that the spatial information bitstream is configured into a attaching frame (e.g., TS bitstream 1506) in a compressed format to attach the attaching frame to the downmix signal bitstream 1502 in the compressed format.

In this case, a TS header 1505 for the TS bitstream 1506 can exist. And, at least one of attaching frame sync information 1507, discriminating information 1508 indicating whether a header of a decoding frame exists within the attaching frame, information for a number of subframes included in the attaching frame and the rest information 1509 can be included in the attaching frame header (e.g., TS header 1505). And, discriminating information indicating whether a start point of the attaching frame and a start point of the decoding frame are matched can be included within the attaching frame.

If the decoding frame header exists within the attaching frame, discriminating information indicating whether there exists position information of a downmix signal to which the spatial information is applied is extracted from the decoding frame header.

Subsequently, the position information of the downmix signal, to which the spatial information is applied, can be extracted according to the discriminating information.

FIG. 16 is a flowchart of a method of encoding a spatial information bitstream embedded in a downmix signal by insertion frames of various sizes according to the present invention.

Referring to FIG. 16, an audio signal is downmixed from a multi-channel audio signal (1601, 1602). In this case, the downmix signal may be a mono, stereo or multi-channel audio signal.

And, spatial information is extracted from the multi-channel audio signal (1601, 1603).

A spatial information bitstream is then generated using the extracted spatial information (1604). The generated spatial information can be embedded in the downmix signal by an insertion frame unit having a length corresponding to an integer multiplication of a decoding frame length per frame.

If a decoding frame length (S) is greater than a insertion frame length (N) (1605), the insertion frame length (N) is configured equal to one S by binding a plurality of Ns together (1607).

If the decoding frame length (S) is smaller than the insertion frame length (N) (1606), the insertion frame length (N) is configured equal to one N by binding a plurality of Ss together (1608).

If the decoding frame length (S) is equal to the insertion frame length (N), the insertion frame length (N) is configured equal to the decoding frame length (S) (1609).

The spatial information bitstream configured in the above-explained manner is embedded in the downmix signal (1610).

Finally, a whole bitstream including the downmix signal having the spatial information bitstream embedded therein is transferred (1611).

Besides, in the present invention, information for an insertion frame length of a spatial information bitstream can be embedded in a whole bitstream.

FIG. 17 is a flowchart of a method of encoding a spatial information bitstream embedded by a fixed length in a downmix signal according to the present invention.

Referring to FIG. 17, an audio signal is downmixed from a multi-channel audio signal (1701, 1702). In this case, the downmix signal may be a mono, stereo or a multi-channel audio signal.

And, spatial information is extracted from the multi-channel audio signal (1701, 1703).

A spatial information bitstream is then generated using the extracted spatial information (1704).

After the spatial information bitstream has been bound into a bitstream having a fixed length (packet unit), e.g., a transport stream (TS) (1705), the spatial information bitstream of the fixed length is embedded in the downmix signal (1706).

Subsequently, a whole bitstream including the downmix signal having the spatial information bitstream embedded therein is transferred (1707).

Besides, in the present invention, an insertion bit length (i.e., K value) of an insertion area, in which the spatial information bitstream is embedded, is obtained using the downmix signal and the spatial information bitstream can be embedded in the insertion area.

FIG. 18 is a diagram of a first method of embedding a spatial information bitstream in an audio signal downmixed on at least one channel according to the present invention.

In case that a downmix signal is configured with at least one channel, spatial information can be regarded as data in common to the at least one channel. So, a method of embedding the spatial information by dispersing the spatial information on the at least one channel is needed.

FIG. 18 shows a method of embedding the spatial information on one channel of the downmix signal having the at least one channel.

Referring to FIG. 18, the spatial information is embedded in K-bits of the downmix signal. In particular, the spatial information is embedded in one channel only but is not embedded in the other channel. And, the K value can differ per block or channel.

As mentioned in the foregoing description, bits corresponding to the K value may correspond to lower bits of the downmix signal, which does not put limitation on the present invention. In this case, the spatial information bitstream can be inserted in one channel in a bit plane order from LSB or in a sample plane order.

FIG. 19 is a diagram of a second method of embedding a spatial information bitstream in an audio signal downmixed on at least one channel according to the present invention. For the convenience of explanation, FIG. 19 shows a downmix signal having two channels, which does not limitation on the present invention.

Referring to FIG. 19, the second method is carried out in a manner of embedding spatial information in a block-n of one channel (e.g., left channel), a block-n of the other channel (e.g., right channel), a block-(n+1) of the former channel (left channel), etc. in turn. In this case, sync information can be embedded in one channel only.

Although a spatial information bitstream can be embedded in a downmix signal per block, it is able to extract the spatial information bitstream per block or frame in a decoding process.

Since signaling characteristics of the two channels of the downmix signal differ from each other, it is able to allocate K values to the two channels differently by finding respective masking thresholds of the two channels separately. In particular, K_1 and K_2, as shown in the drawing, can be allocated to the two channels, respectively.

In this case, the spatial information can be embedded in each of the channels in a bit plane order from LSB or in a sample plane order.

FIG. 20 is a diagram of a third method of embedding a spatial information bitstream in an audio signal downmixed on at least one channel according to the present invention. FIG. 20 shows a downmix signal having two channels, which does not put limitation on the present invention.

Referring to FIG. 20, the third method is carried out in a manner of embedding spatial information by dispersing it on two channels. In particular, the spatial information is embedded in a manner of alternating a corresponding embedding order for the two channels by sample unit.

Since signaling characteristics of the two channels of the downmix signal differ from each other, it is able to allocate K values to the two channels differently by finding respective masking thresholds of the two channels separately. In particular, K_1 and K_2, as shown in the drawing, can be allocated to the two channels, respectively.

The K values may differ from each other per block. For instance, the spatial information is put in lower K_1 bits of a sample-1 of one channel (e.g., left channel), lower K_2 bits of a sample-1 of the other channel (e.g., right channel), lower K_1 bits of a sample-2 of the former channel (e.g., left channel) and lower K_2 bits of a sample-

2 of the latter channel (e.g., right channel), in turn.

In the drawing, a numeral within parentheses indicates an order of filling the spatial information bitstream. Although FIG. 20 shows that the spatial information bitstream is filled from MSB, the spatial information bitstream can be filled from LSB.

FIG. 21 is a diagram of a fourth method of embedding a spatial information bitstream in an audio signal downmixed on at least one channel according to the present invention. FIG. 21 shows a downmix signal having two channels, which does not put limitation on the present invention.

Referring to FIG. 21, the fourth method is carried out in a manner of embedding spatial information by dispersing it on at least one channel. In particular, the spatial information is embedded in a manner of alternating a corresponding embedding order for two channels by bit plane unit from LSB.

Since signaling characteristics of the two channels of the downmix signal differ from each other, it is able to allocate K values (K_1 and K_2) to the two channels differently by finding respective masking thresholds of the two channels separately. In particular, K_1 and K_2, as shown in the drawing, can be allocated to the two channels, respectively.

The K values may differ from each other per block. For instance, the spatial information is put in a least significant 1 bit of a sample-1 of one channel (e.g., left channel), a least significant 1 bit of a sample-1 of the other channel (e.g., right channel), a least significant 1 bit of a sample-2 of the former channel (e.g., left channel) and a least significant 1 bit of a sample-2 of the latter channel (e.g., right channel), in turn. In the drawing, a numeral within a block indicates an order of filling spatial information.

In case that an audio signal is stored in a storage medium (e.g., stereo CD) having no auxiliary data area or is transferred by SPDIF or the like, L/R channel is interleaved by sample unit. So, it is advantageous for a decoder to process a audio signal according to a received order if the audio signal is stored by the third or fourth method.

And, the fourth method is applicable to a case that a spatial information bitstream is stored by being rearranged by bit plane unit.

As mentioned in the foregoing description, in case that a spatial information bitstream is embedded by being dispersed on two channels, it is able to differently allocate K values to the channels, respectively. In this case, it is possible to separately transfer the K value per each of the channels within the bitstream. In case that a plurality of K values are transferred, differential encoding is applicable to a case of encoding the K values.

FIG. 22 is a diagram of a fifth method of embedding a spatial information bitstream in an audio signal downmixed on at least one channel according to the present invention. FIG. 22 shows a downmix signal having two channels, which does not put limitation on the present invention.

Referring to FIG. 22, the fifth method is carried out in a manner of embedding spatial information by dispering it on two channels. In particular, the fifth method is carried out in a manner of inserting the same value in each of the two channels repeatedly.

In this case, a value of the same sign can be inserted in each of the at least two channels or the values differing in signs can be inserted in the at least two channels, respectively.

For instance, a value of 1 is inserted in each of the two channels or values of 1 and -1 can be alternately inserted in the two channels, respectively.

The fifth method is advantageous in facilitating a transmission error to be checked by comparing a least significant insertion bits (e.g., K bits) of at least one channel.

In particular, in case of transferring a mono audio signal to a stereo medium such as a CD, since channel-L (left channel) and channel-R (right channel) of a downmix signal are identical to each other, robustness and the like can be enhanced by equalizing the inserted spatial information. In this case, the spatial information can be embedded in each of the channels in a bit plane order from LSB or in a sample plane order.

FIG. 23 is a diagram of a sixth method of embedding a spatial information bitstream in an audio signal downmixed on at least one channel according to the present invention.

The sixth method relates to a method of inserting spatial information in a downmix signal having at least one channel in case that a frame of each channel includes a plurality of blocks (length B).

Referring to FIG. 23, insertion bit lengths (i.e., K values) may have different values per channel and block, respectively or may have the same value per channel and block.

The insertion bit lengths (e.g., K_1, K_2, K_3 and K_4) can be stored within a frame header transmitted once for a whole frame. And, the frame header cab be located at LSB.

In this case, the header can be inserted by bit plane unit. And, spatial information data can be alternately inserted by sample unit or by block unit. In FIG. 23, a number of blocks within a frame is 2. So, a length (B) of the block is N/2. In this case, a number of bits inserted in the frame is (K1+K2+K3+K4)*B.

FIG. 24 is a diagram of a seventh method of embedding a spatial information bitstream in an audio signal downmixed on at least one channel according to the present invention. FIG. 24 shows a downmix signal having two channels, which does not put limitation on the present invention.

Referring to FIG. 22, the seventh method is carried out in a manner of embedding spatial information by dispersing it on two channels. In particular, the seventh method is characterized in mixing a method of inserting the spatial information in the two channels in a bit plane order from LSB or MSB alternately and a method of inserting the spatial information in the two channels alternately by sample plane order.

The method is performed by frame unit or can be performed by block unit.

Hatching portions 1 to C, as shown in FIG. 24, correspond to a header and can be inserted in LSB or MSB in a bit plane order to facilitate a search for an insertion frame sync word.

Other portions (non-hatching portions) C+1 and higher correspond to portions excluding the header and can be inserted in two channels alternately by sample unit to facilitate spatial information data to be extracted out. Insertion bit sizes (e.g., K values) can have different or same values from each other per channel and block. And, the all insertion bit lengths can be included in the header.

FIG. 25 is a flowchart of a method of encoding spatial information to be embedded in a downmix signal having at least one channel according to the present invention.

Referring to FIG. 25, an audio signal is downmixed into one channel from a multi-channel audio signal (2501, 2502). And, spatial information is extracted from the multi-channel audio signal (2501, 2503).

A spatial information bitstream is then generated using the extracted spatial information (2504).

The spatial information bitstream is embedded in the downmix signal having the at least one channel (2505). In this case, one of the seven methods for embedding the spatial information bitstream in the at least one channel can be used.

Subsequently, a whole stream including the downmix signal having the spatial information bitstream embedded therein is transferred (2506). In this case, the present invention finds a K value using the down mix signal and can embed the spatial information bitstream in the K bits.

FIG. 26 is a flowchart of a method of decoding a spatial information bitstream embedded in a downmix signal having at least one channel according to the present invention.

Referring to FIG. 26, a spatial decoder receives a bitstream including a downmix signal in which a spatial information bitstream is embedded (2601).

The downmix signal is detected from the received bitstream (2602).

The spatial information bitstream embedded in the downmix signal having the at least one channel is extracted and decoded from the received bitstream (2603).

Subsequently, the downmix signal is converted to a multi-channel signal using the spatial information obtained from the decoding (2604).

The present invention extracts discriminating information for an order of embedding the spatial information bitstream and can extract and decode the spatial information bitstream using the discriminating information.

And, the present invention extracts information for a K value from the spatial information bitstream and can decode the spatial information bitstream using the K value.

INDUSTRIAL APPLICABILITY

Accordingly, the present invention provides the following effects or advantages.

First of all, in coding a multi-channel audio signal according to the present invention, spatial information is embedded in a downmix signal. Hence, a multi-channel audio signal can be stored/reproduced in/from a storage medium (e.g., stereo CD) having no auxiliary data area or an audio format having no auxiliary data area.

Secondly, spatial information can be embedded in a downmix signal by various frame lengths or a fixed frame length. And, the spatial information can be embedded in a downmix signal having at least one channel. Hence, the present invention enhances encoding and decoding efficiencies.

While the present invention has been described and illustrated herein with reference to the preferred embodiments thereof, it will be apparent to those skilled in the art that various modifications and variations can be made therein without departing from the spirit and scope of the invention. Thus, it is intended that the present invention covers the modifications and variations of this invention that come within the scope of the appended claims and their equivalents.

WHAT IS CLAIMED IS:

1. A method of decoding an audio signal, comprising :

extracting side information embedded in the audio signal, wherein the side information is dispersed corresponding to at least one channel of the audio signal; and

decoding the audio signal using the side information.

2. The method of claim 1, wherein the side information is embedded in an insertion area of the audio signal by a block unit.

3. The method of claim 2, wherein the side information in the insertion area is embedded in a sample plane order or a bit plane order.

4. The method of claim 3, wherein the side information in the insertion area is embedded from MSB (most significant bit) or LSB (least significant bit).

5. The method of claim 2, wherein the side information is embedded in the insertion area by

alternating channels.

6. The method of claim 1, further comprising extracting sync information for the side information from one channel of the audio signal.

7. The method of claim 1, the extracting of side information comprising extracting the side information by a sample unit up to an insertion frame end of the side information.

8. The method of claim 8, wherein the side information is repeatedly embedded in the audio signal having at least two channels with a same value or with values of opposite signs.

9. The method of claim 1, wherein a header for the side information is embedded in the audio signal having at least one channel in a bit plane order and wherein an area except the header is embedded in a sample plane order.

10. The method of claim 1, further comprising extracting an insertion bit length for the side information from a header of the side information.

11. The method of claim 1, wherein the audio signal includes a downmix signal for a multi-channel signal.

12. The method of claim 1, wherein the side information includes spatial information for a multi-channel signal.

13. A method of encoding an audio signal, comprising :

generating side information necessary for decoding the audio signal; and

embedding the side information in the audio signal having at least one channel by dispersing the side information.

14. The method of claim 13, the embedding of the side information comprising embedding the side information in the audio signal having a plurality of blocks by a block unit.

15. The method of claim 14, wherein an insertion bit length of the embedded side information is obtained per the block.

16. The method of claim 14, wherein the insertion bit length of the embedded side information is obtained per a channel of the audio signal.

17. The method of claim 13, the embedding of the side information further comprising, if a number of bits required to embed the side information in a frame of audio signal is smaller than a number of bits allowed to embed the side information of the audio signal, replacing remaining bits per an insertion frame with zeros, a random signal, an original audio signal, a tail sequence or a combination of at least two of the zeros, the random signal, the original audio signal and the tail sequence.

18. A data structure comprising:

an audio signal; and

side information necessary for decoding the audio signal embedded in non-recognizable components of the audio signal having at least one channel by being dispersed.

19. The data structure of claim 18, wherein the side information is embedded in the audio signal in a sample plane order.

20. The data structure of claim 18, wherein the side information is embedded in the audio signal in a bit plane order.

21. The data structure of claim 18, wherein a header for the side information is embedded in the audio signal in a bit plane order and wherein an area except the header is embedded in a sample plane order.

22. An apparatus for encoding an audio signal, comprising:

a side information generating unit for generating side information necessary for decoding an audio signal; and

an embedding unit for embedding the side information in the audio signal having at least one channel by dispersing the side information.

23. An apparatus for decoding an audio signal, comprising:

an embedded signal decoding unit for extracting side information embedded in the audio signal having at least one channel by being dispersed; and a multi-channel generating unit for decoding the audio signal using the additional information.

FIG. 1

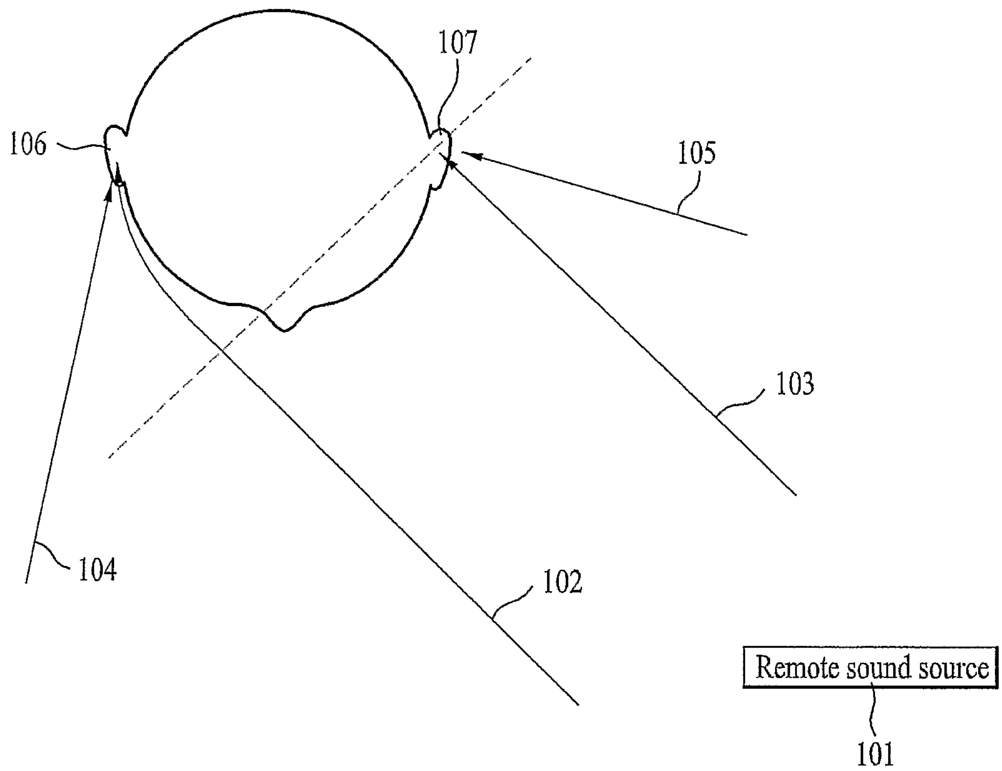


FIG. 2

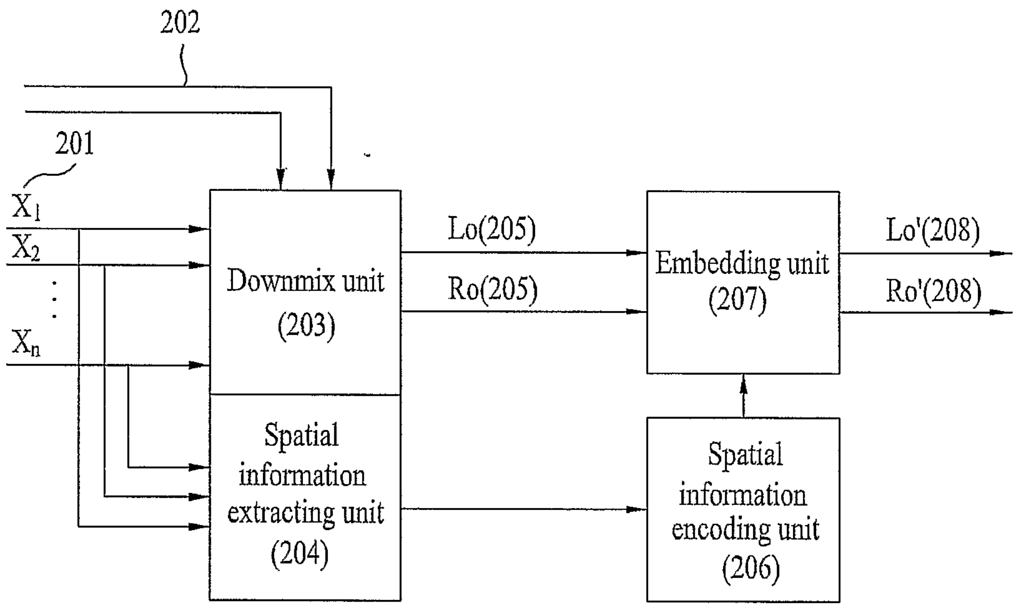

202
201
X1
X2
Xn
Downmix unit (203)
Spatial information extracting unit (204)
Lo(205)
Ro(205)
Embedding unit (207)
Spatial information encoding unit (206)
Lo'(208)
Ro'(208)

FIG. 3

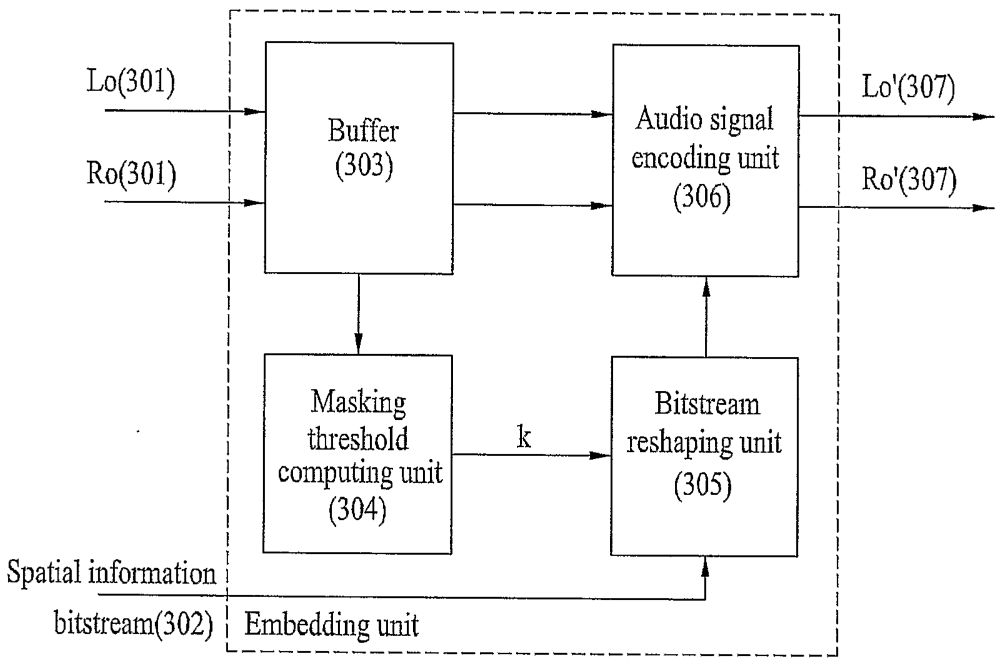
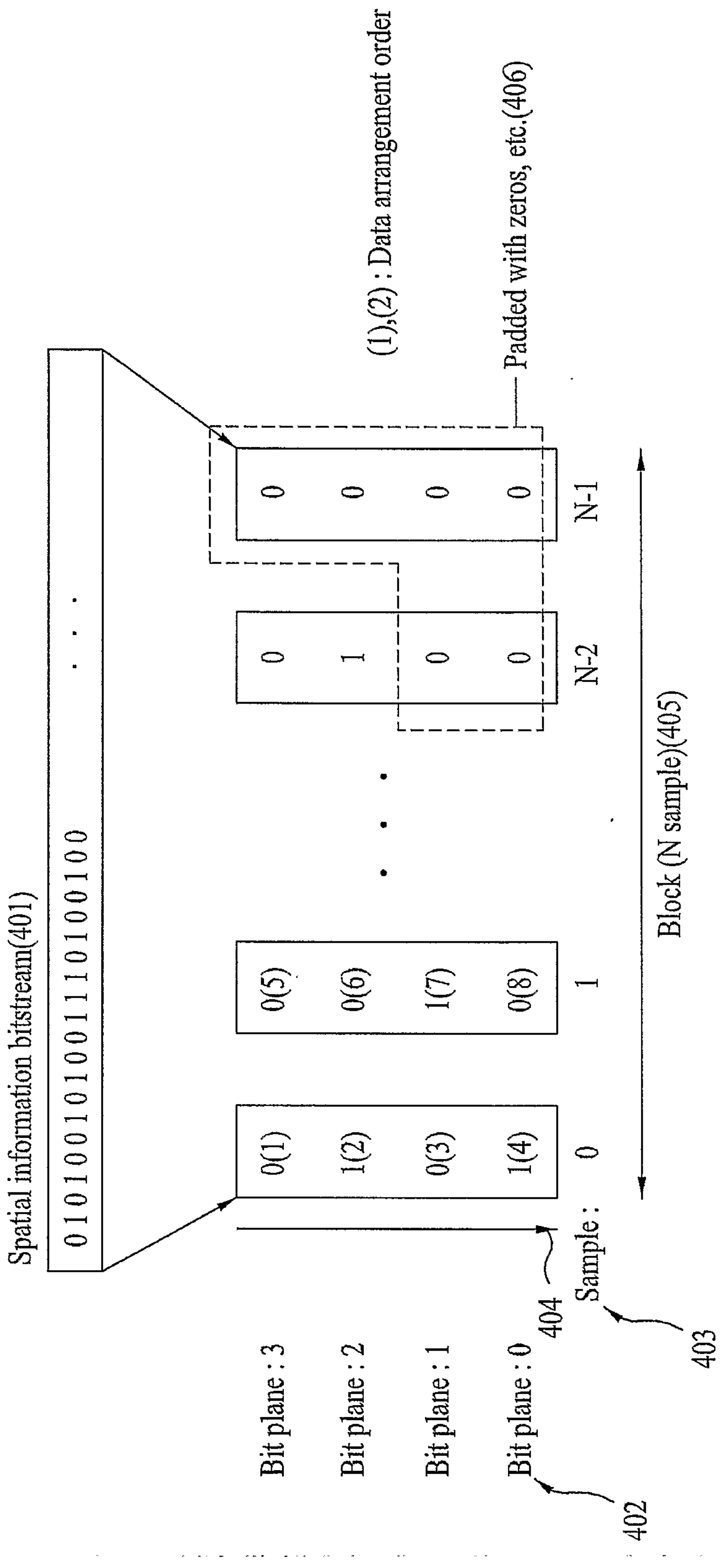

Lo(301)
Ro(301)
Buffer (303)
Audio signal encoding unit (306)
Lo'(307)
Ro'(307)
Masking threshold computing unit (304)
k
Bitstream reshaping unit (305)
Spatial information bitstream(302)
Embedding unit

FIG. 4



Spatial information bitstream(401)
0 1 0 1 0 0 1 0 1 0 0 1 1 1 0 1 0 0 1 0 0 . . .
(1),(2) : Data arrangement order
Padded with zeros, etc.(406)
Bit plane : 3
Bit plane : 2
Bit plane : 1
Bit plane : 0
0(1)
1(2)
0(3)
1(4)
0(5)
0(6)
1(7)
0(8)
0
1
0
0
0
0
0
0
Sample :
0
1
N-2
N-1
Block (N sample)(405)
402
403
404

FIG. 5

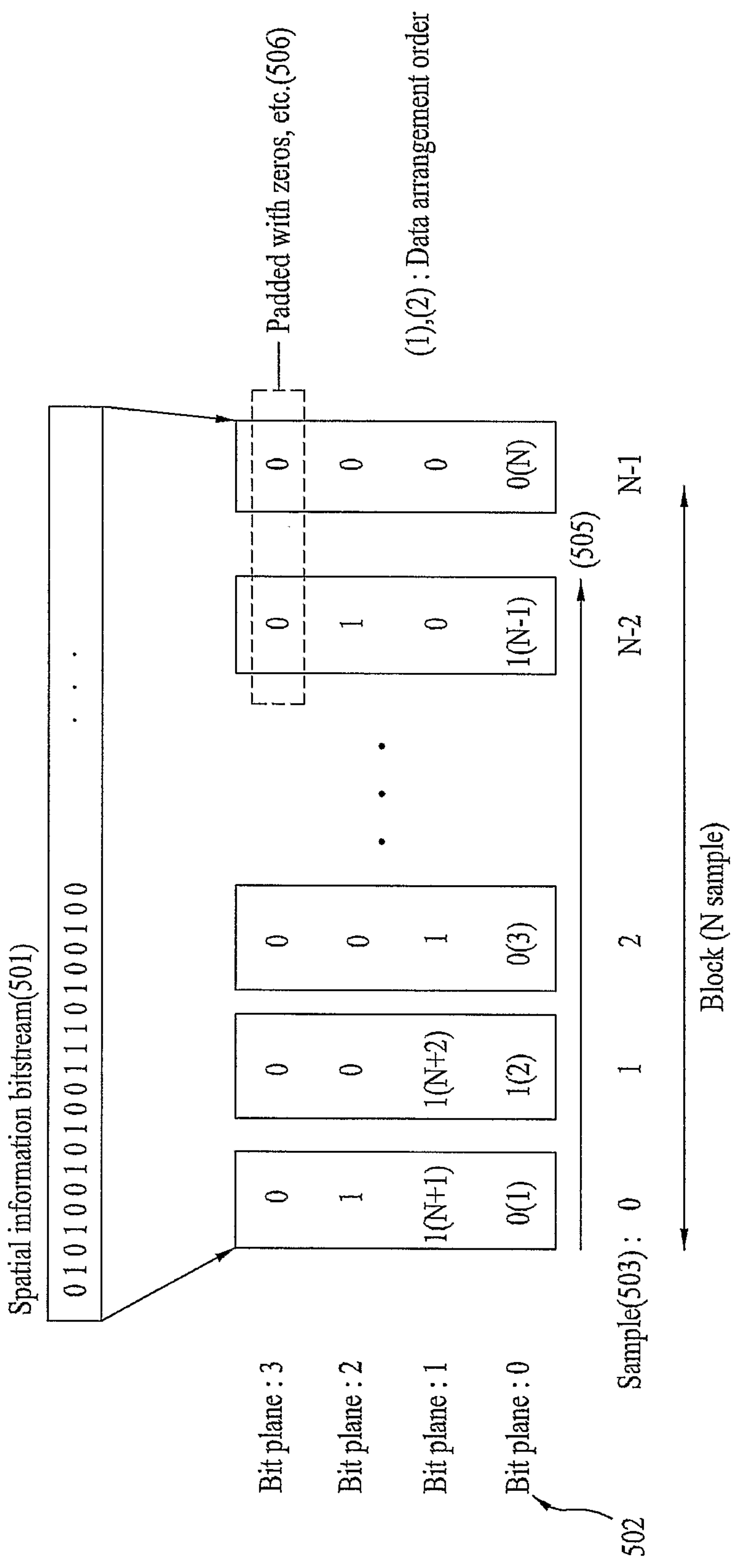

Spatial information bitstream(501)
0 1 0 1 0 0 1 0 1 0 1 0 0 1 1 1 0 1 0 1 0 0 1 0 0
Padded with zeros, etc.(506)
(1),(2) : Data arrangement order
Bit plane : 3
Bit plane : 2
Bit plane : 1
Bit plane : 0
502
0
1
1(N+1)
0(1)
0
0
1(N+2)
1(2)
0
0
1
0(3)
0
1
0
1(N-1)
0
0
0
0(N)
(505)
Sample(503) : 0
1
2
N-2
N-1
Block (N sample)

FIG. 6A

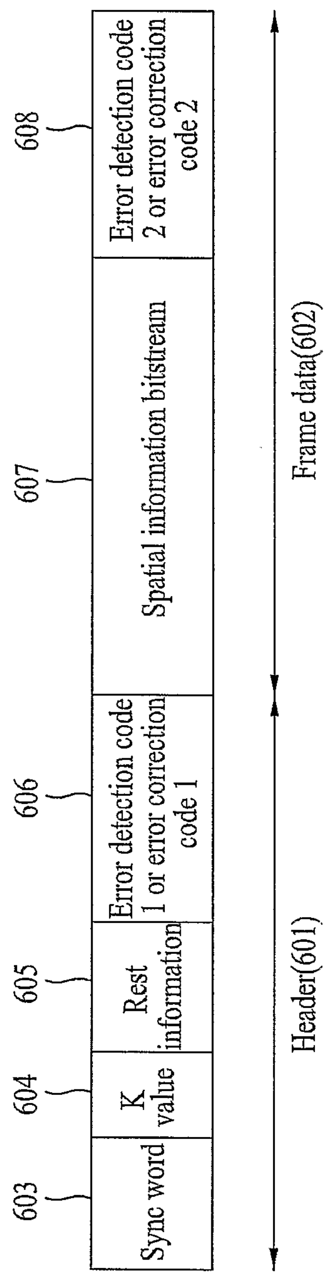

603
604
605
606
607
608
Sync word
K value
Rest information
Error detection code 1 or error correction code 1
Spatial information bitstream
Error detection code 2 or error correction code 2
Header(601)
Frame data(602)

FIG. 6B

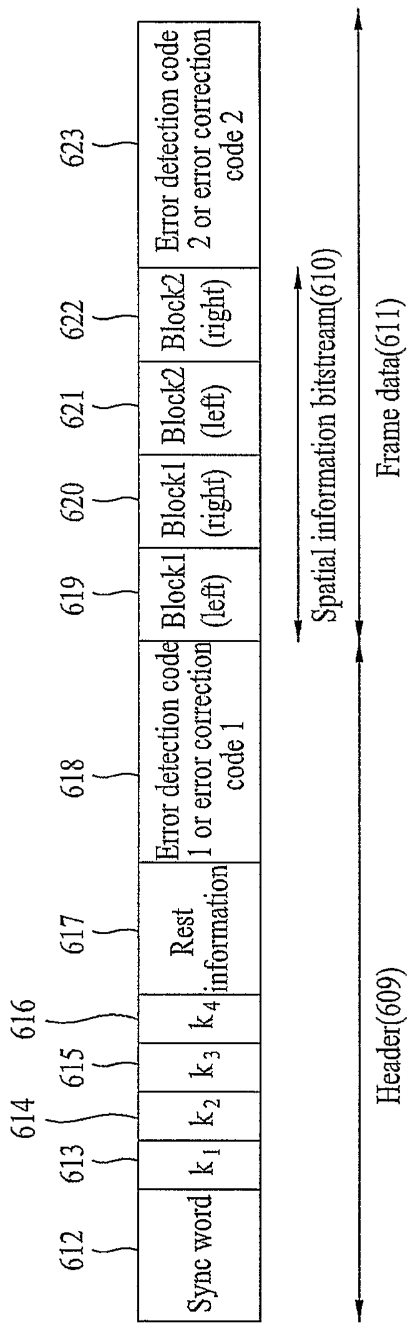

612
613
614
615
616
617
618
619
620
621
622
623
Sync word
k_1
k_2
k_3
k_4
Rest information
Error detection code 1 or error correction code 1
Block1 (left)
Block1 (right)
Block2 (left)
Block2 (right)
Error detection code 2 or error correction code 2
Spatial information bitstream(610)
Header(609)
Frame data(611)

FIG. 7

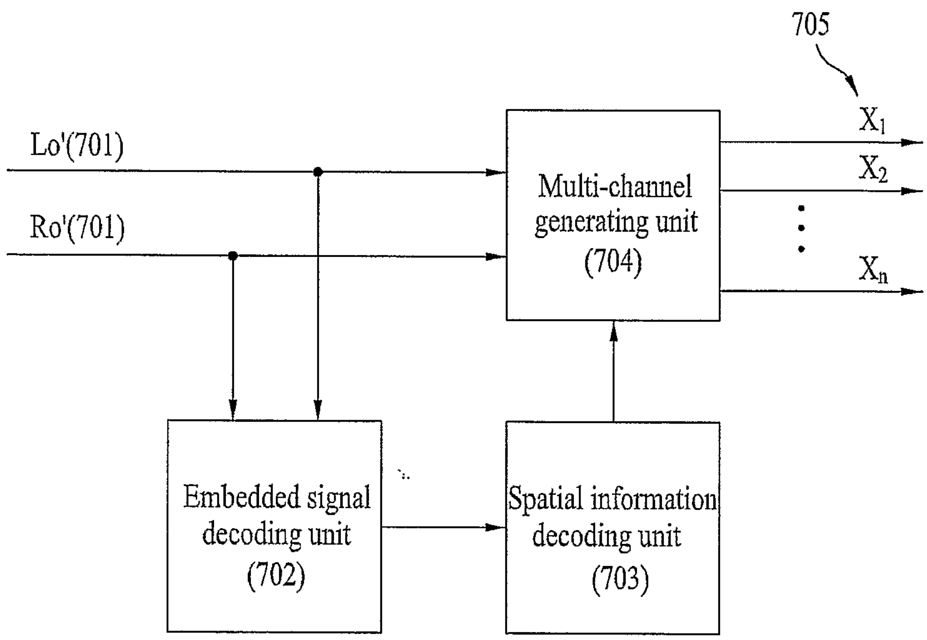

705
Lo'(701)
Ro'(701)
Multi-channel generating unit (704)
X1
X2
Xn
Embedded signal decoding unit (702)
Spatial information decoding unit (703)

FIG. 8

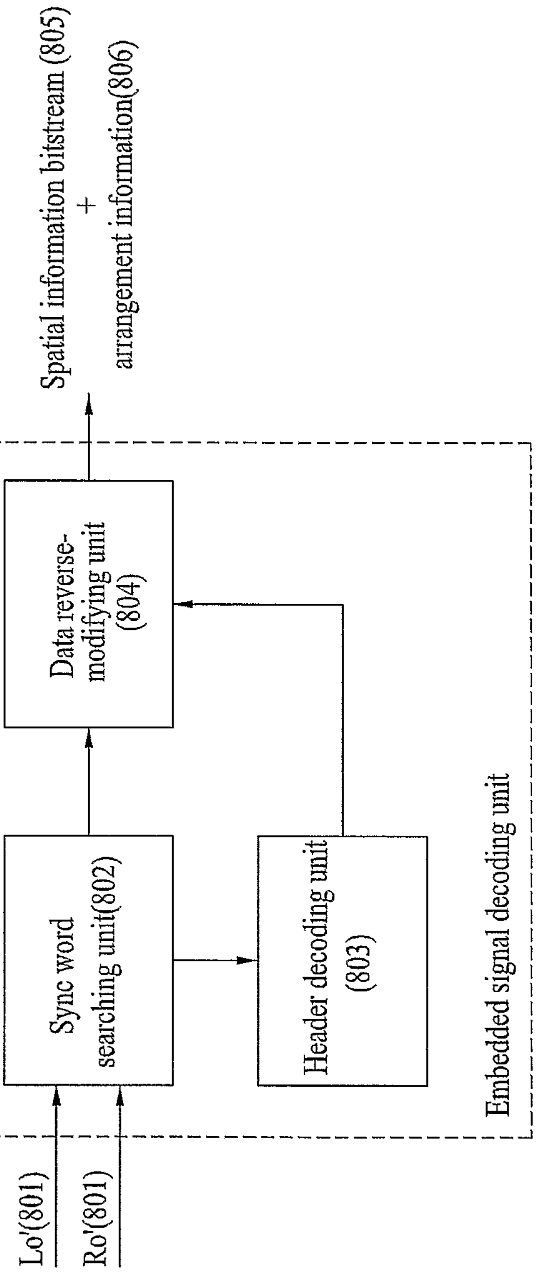

Spatial information bitstream (805)
+
arrangement information(806)
Data reverse-modifying unit (804)
Sync word searching unit(802)
Header decoding unit (803)
Embedded signal decoding unit
Lo'(801)
Ro'(801)

FIG. 9

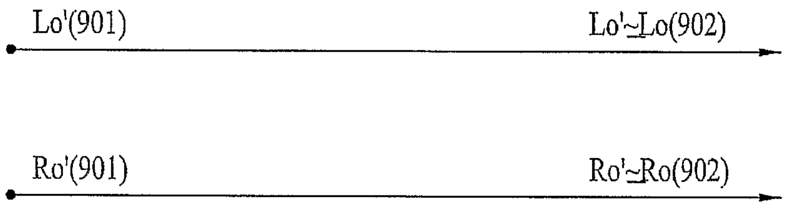


FIG. 10

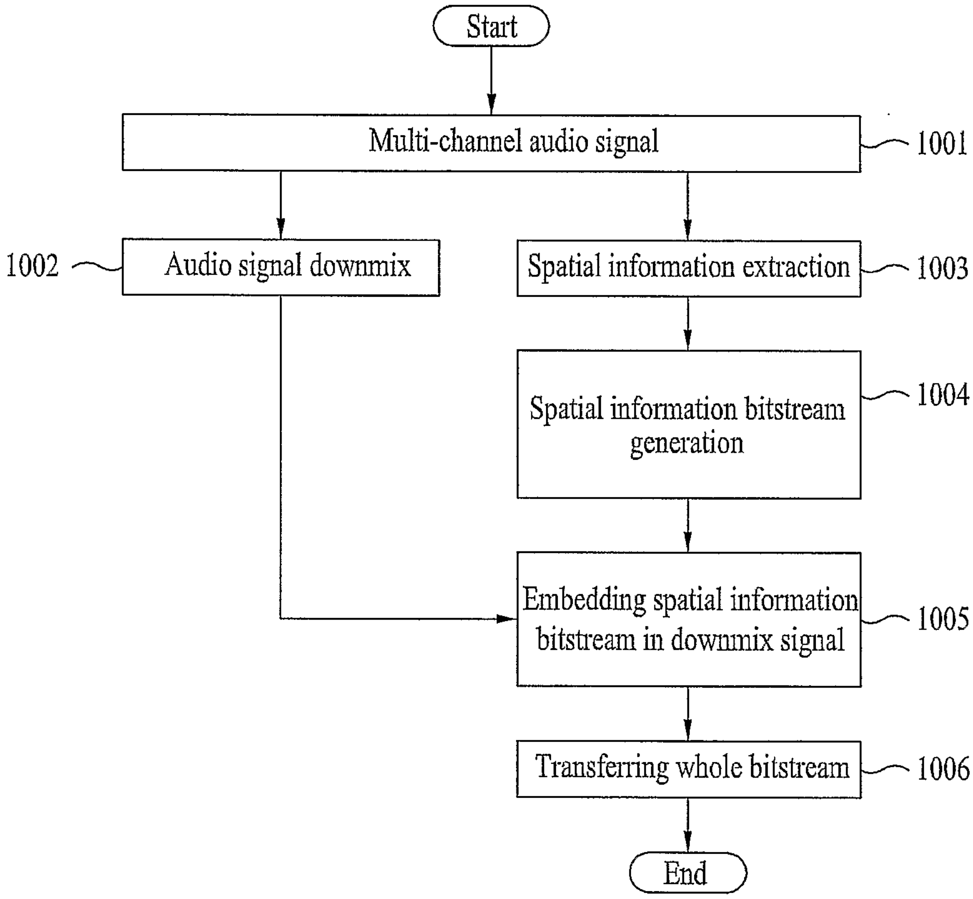

Start
Multi-channel audio signal
1001
1002
Audio signal downmix
Spatial information extraction
1003
Spatial information bitstream generation
1004
Embedding spatial information bitstream in downmix signal
1005
Transferring whole bitstream
1006
End

FIG. 11

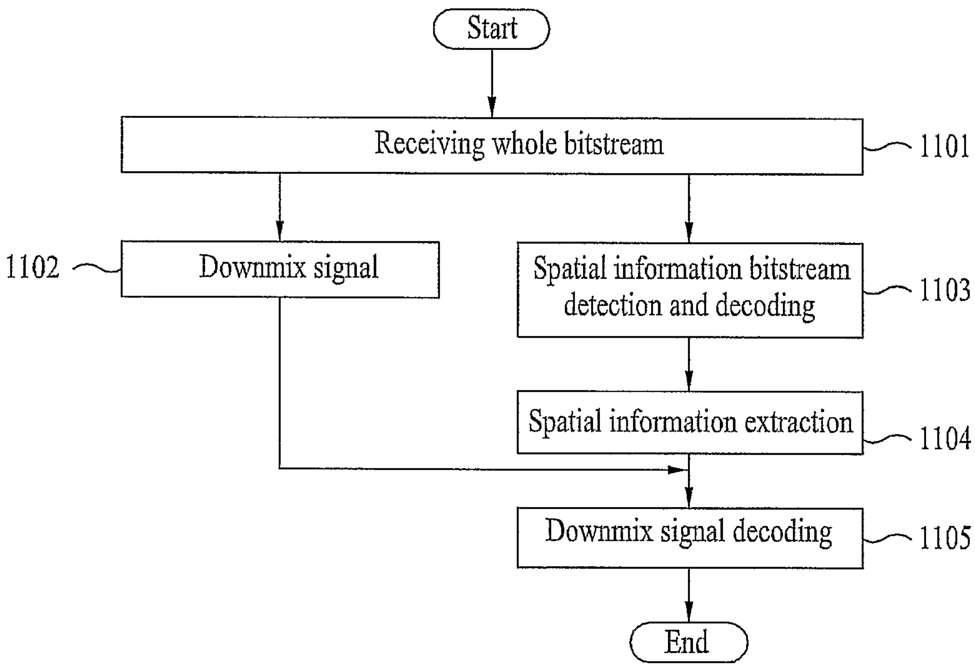

Start
Receiving whole bitstream
1101
1102
Downmix signal
Spatial information bitstream detection and decoding
1103
Spatial information extraction
1104
Downmix signal decoding
1105
End

FIG. 12

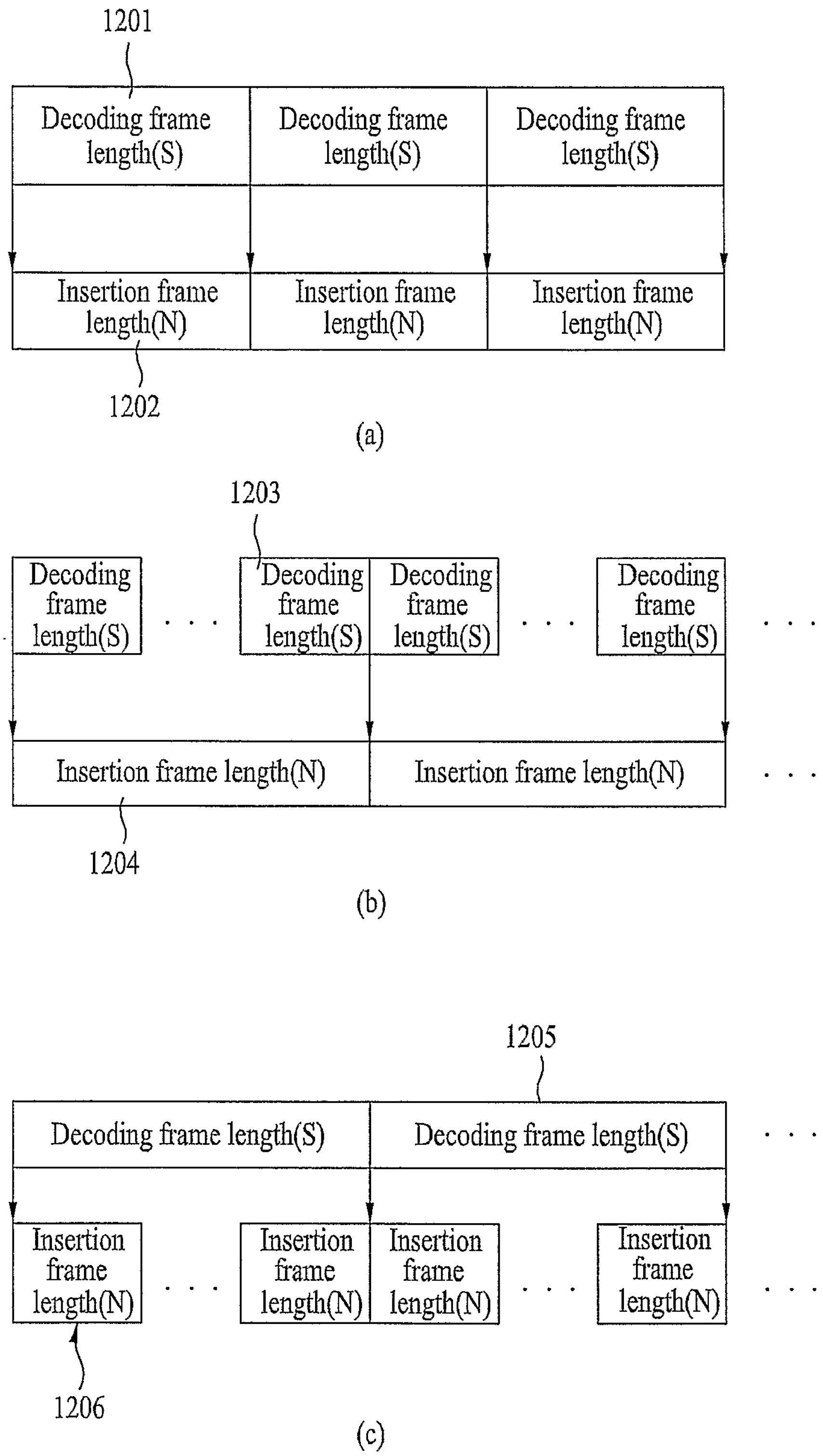

1201
Decoding frame length(S)
Decoding frame length(S)
Decoding frame length(S)
Insertion frame length(N)
Insertion frame length(N)
Insertion frame length(N)
1202
(a)
1203
Decoding frame length(S)
Decoding frame length(S)
Decoding frame length(S)
Decoding frame length(S)
Insertion frame length(N)
Insertion frame length(N)
1204
(b)
1205
Decoding frame length(S)
Decoding frame length(S)
Insertion frame length(N)
Insertion frame length(N)
Insertion frame length(N)
Insertion frame length(N)
1206
(c)

FIG. 13

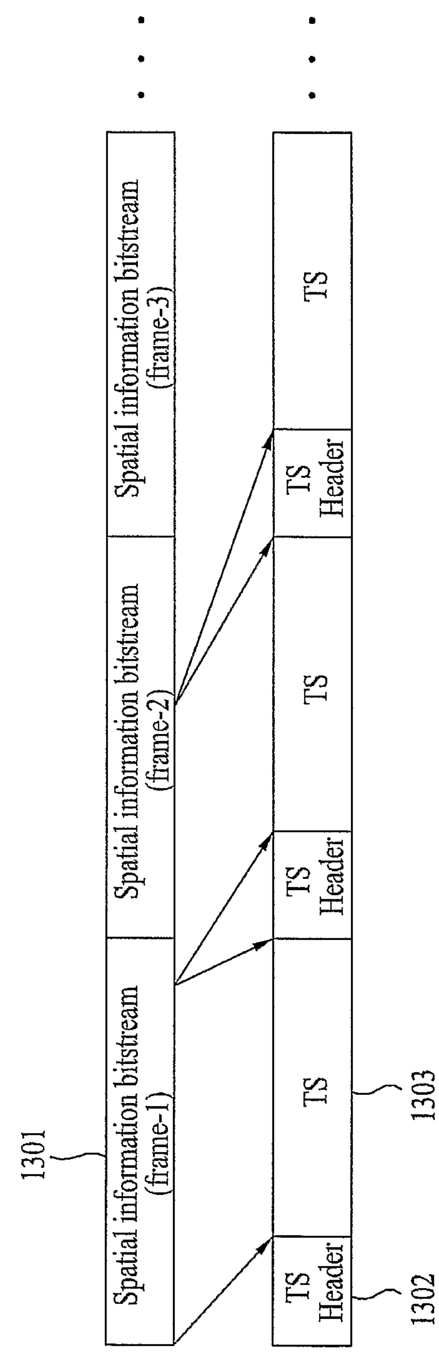

Spatial information bitstream (frame-1)
Spatial information bitstream (frame-2)
Spatial information bitstream (frame-3)
TS Header
TS
TS Header
TS
TS Header
TS
1301
1302
1303

FIG. 14A

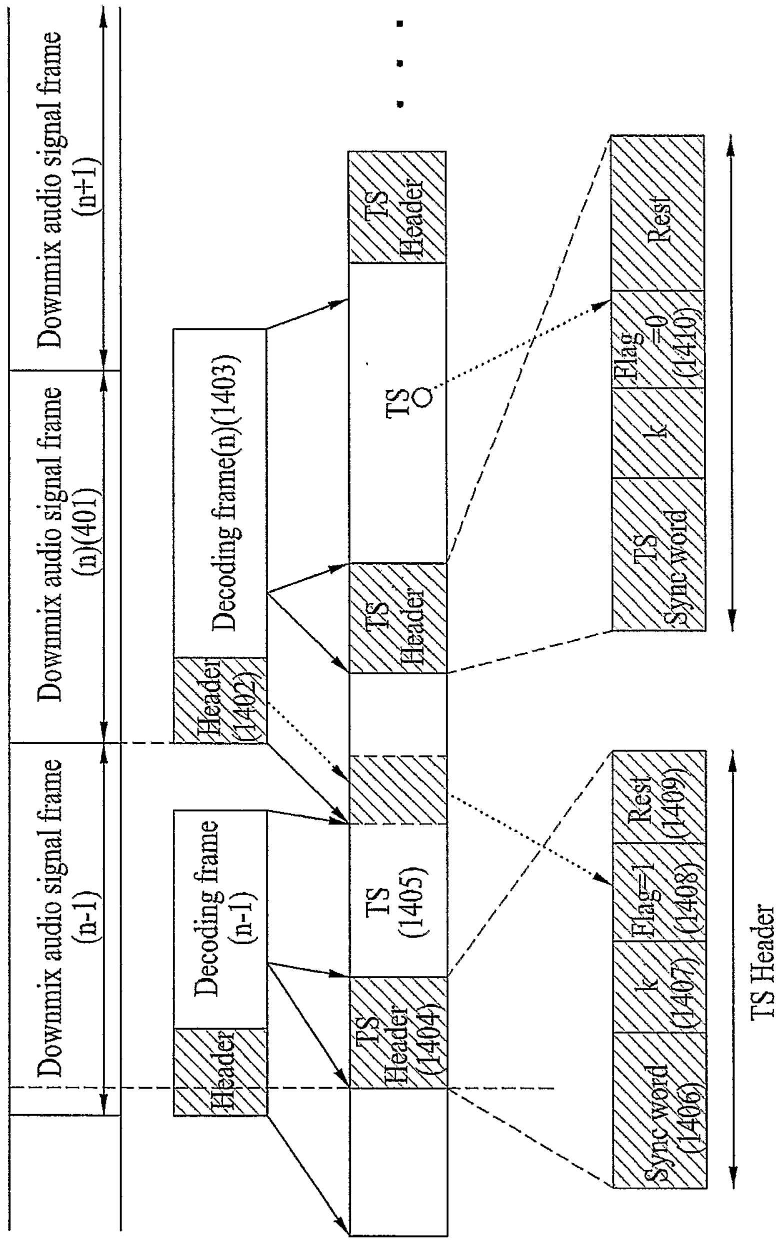
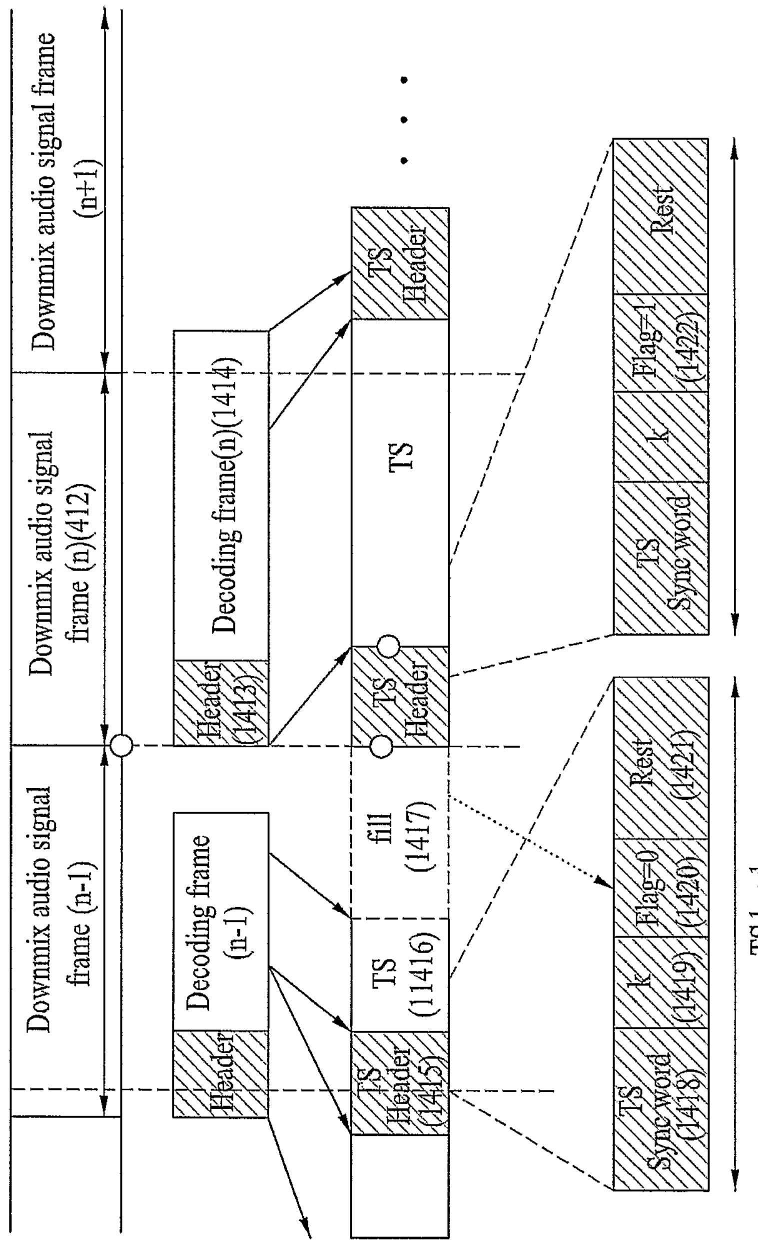

Downmix audio signal frame (n-1)
Downmix audio signal frame (n)(401)
Downmix audio signal frame (n+1)
Header
Decoding frame (n-1)
Header (1402)
Decoding frame(n)(1403)
TS Header (1404)
TS (1405)
TS Header
TS
TS Header
Sync word (1406)
k (1407)
Flag=1 (1408)
Rest (1409)
TS Header
TS Sync word
k
Flag =0 (1410)
Rest

FIG. 14B



Downmix audio signal frame (n-1)
Downmix audio signal frame (n)(412)
Downmix audio signal frame (n+1)
Header
Decoding frame (n-1)
Header (1413)
Decoding frame(n)(1414)
TS Header (1415)
TS (11416)
fill (1417)
TS Header
TS
TS Header
TS Sync word (1418)
K (1419)
Flag=0 (1420)
Rest (1421)
TS Sync word
K
Flag=1 (1422)
Rest
TS header

FIG. 15

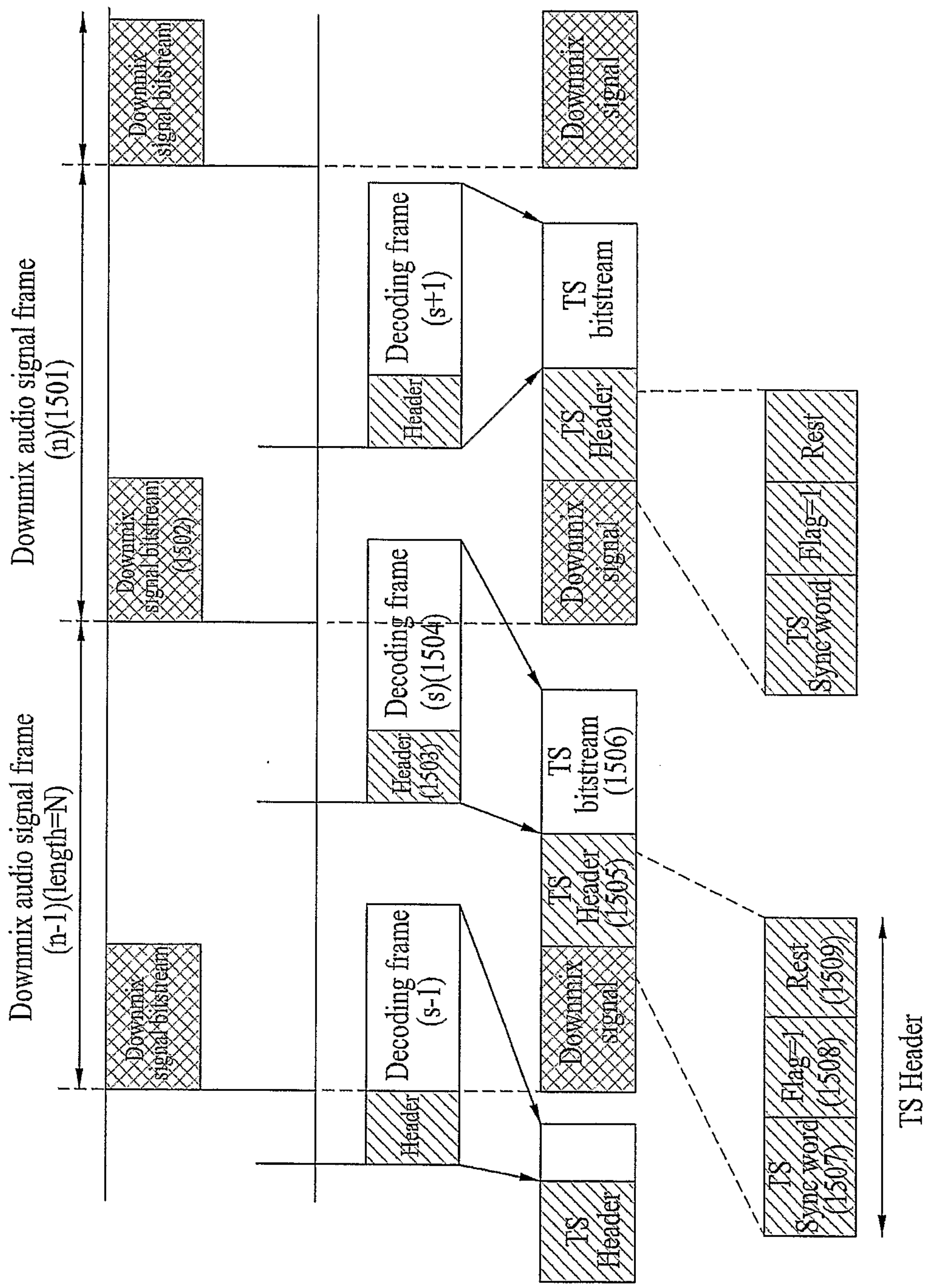
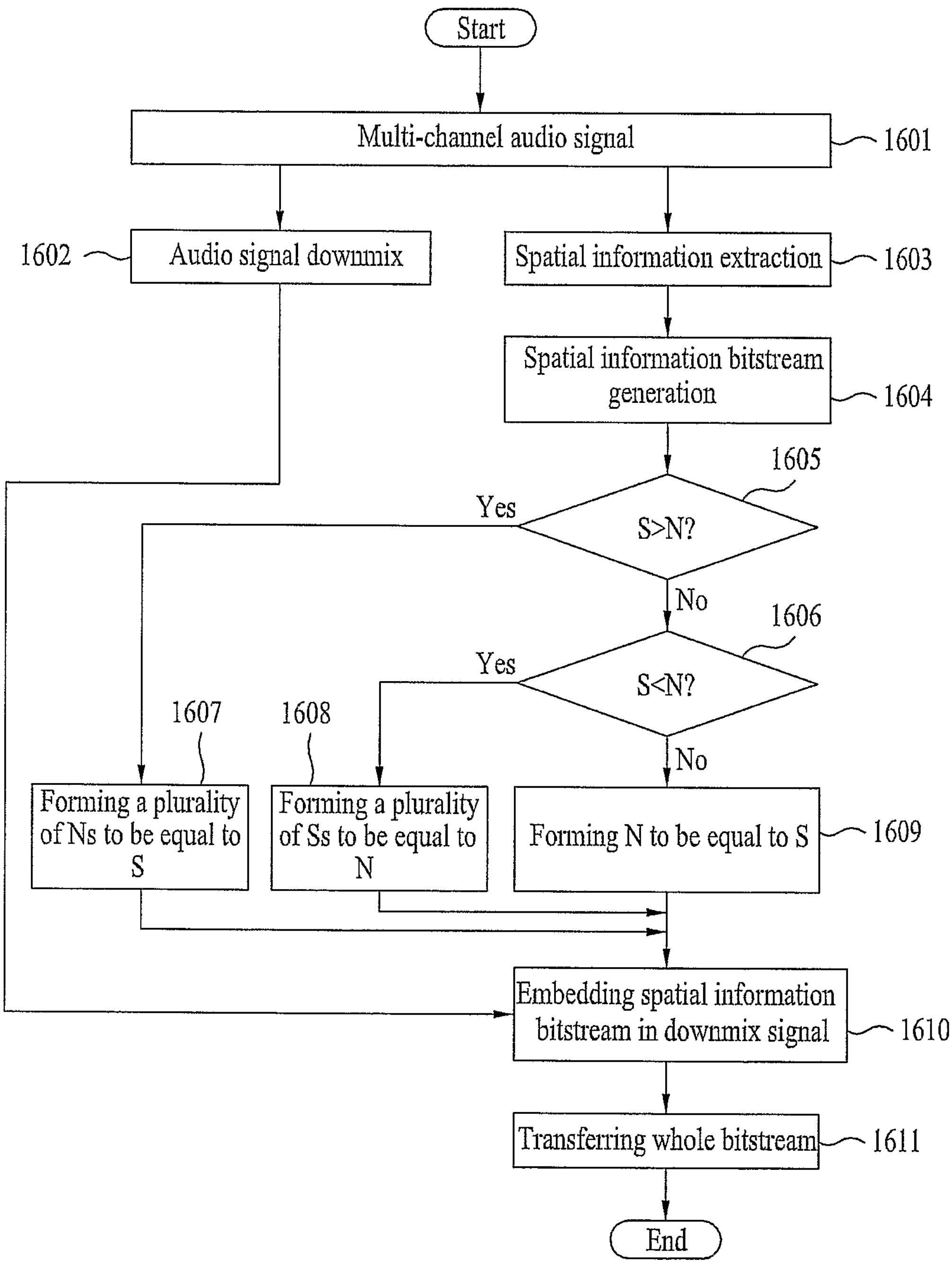

Downmix audio signal frame
(n-1)(length=N)
Downmix audio signal frame
(n)(1501)
Downmix signal bitstream
Downmix signal bitstream (1502)
Downmix signal bitstream
Header
Decoding frame
(s-1)
Header (1503)
Decoding frame
(s)(1504)
Header
Decoding frame
(s+1)
TS Header
Downmix signal
TS Header (1505)
TS bitstream (1506)
Downmix signal
TS Header
TS bitstream
Downmix signal
TS Sync word (1507)
Flag=1 (1508)
Rest (1509)
TS Header
TS Sync word
Flag=1
Rest

FIG. 16



Start
Multi-channel audio signal
1601
1602
Audio signal downmix
Spatial information extraction
1603
Spatial information bitstream generation
1604
1605
Yes
S>N?
No
1606
Yes
S<N?
No
1607
1608
Forming a plurality of Ns to be equal to S
Forming a plurality of Ss to be equal to N
Forming N to be equal to S
1609
Embedding spatial information bitstream in downmix signal
1610
Transferring whole bitstream
1611
End

FIG. 17

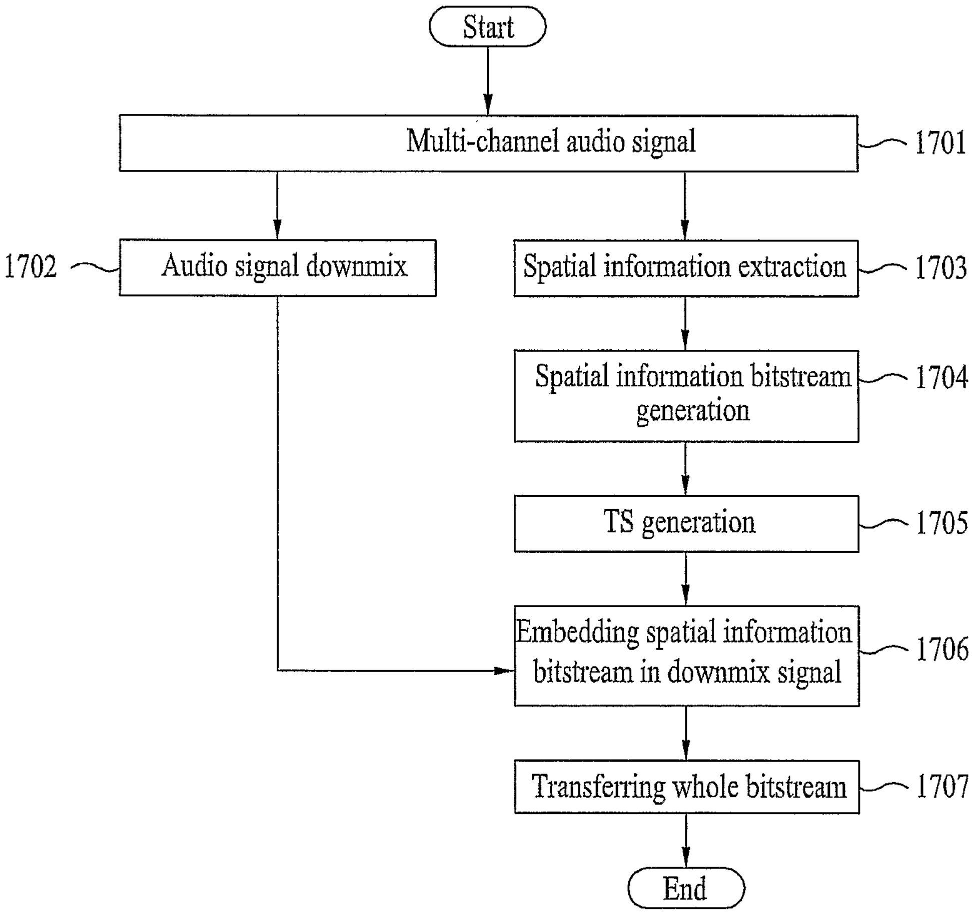

Start
Multi-channel audio signal
1701
1702
Audio signal downmix
Spatial information extraction
1703
Spatial information bitstream generation
1704
TS generation
1705
Embedding spatial information bitstream in downmix signal
1706
Transferring whole bitstream
1707
End

FIG. 18

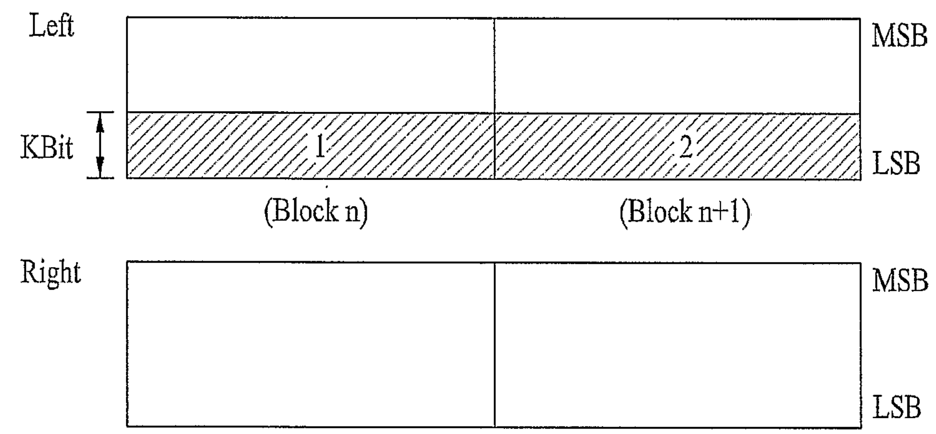

Left
MSB
KBit
1
2
LSB
(Block n)
(Block n+1)
Right
MSB
LSB

FIG. 19

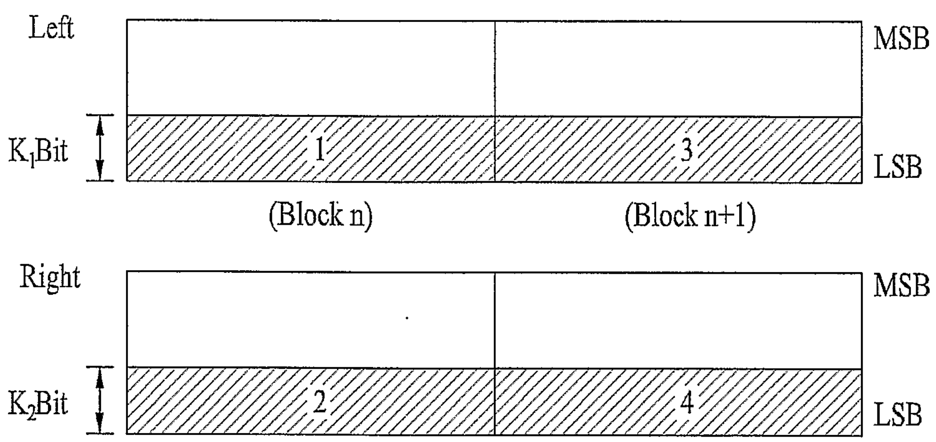

Left
MSB
K_1Bit
1
3
LSB
(Block n)
(Block n+1)
Right
MSB
K_2Bit
2
4
LSB

FIG. 20

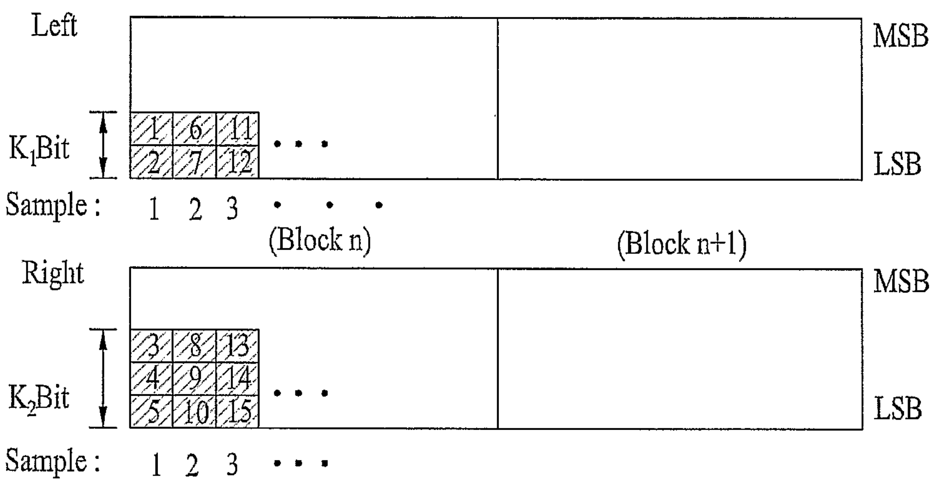

Left
MSB
1 6 11
2 7 12
K1Bit
LSB
Sample : 1 2 3
(Block n)
(Block n+1)
Right
MSB
3 8 13
4 9 14
5 10 15
K2Bit
LSB
Sample : 1 2 3

FIG. 21

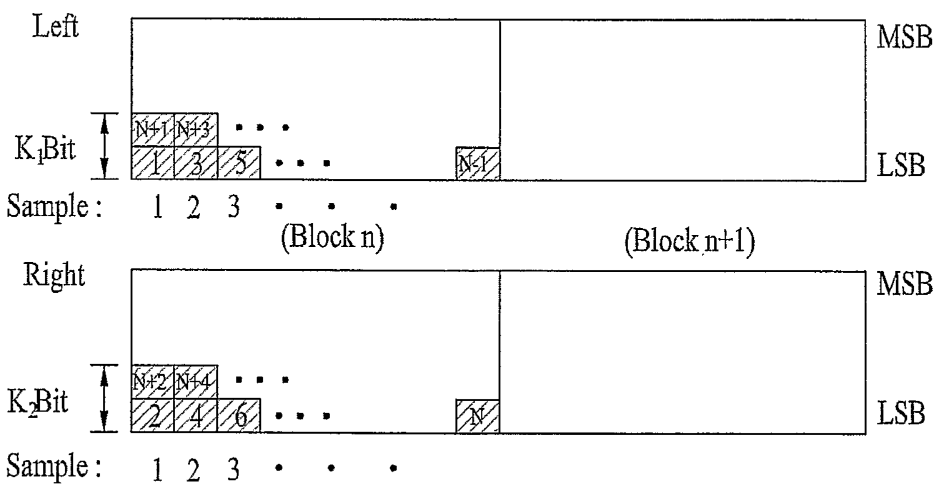

Left
MSB
N+1 N+3
1 3 5
N-1
K1Bit
LSB
Sample : 1 2 3
(Block n)
(Block n+1)
Right
MSB
N+2 N+4
2 4 6
N
K2Bit
LSB
Sample : 1 2 3

FIG. 22

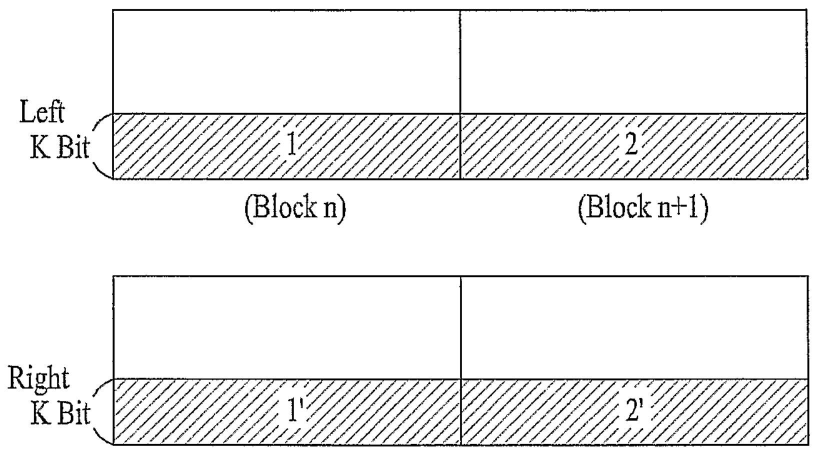


FIG. 23

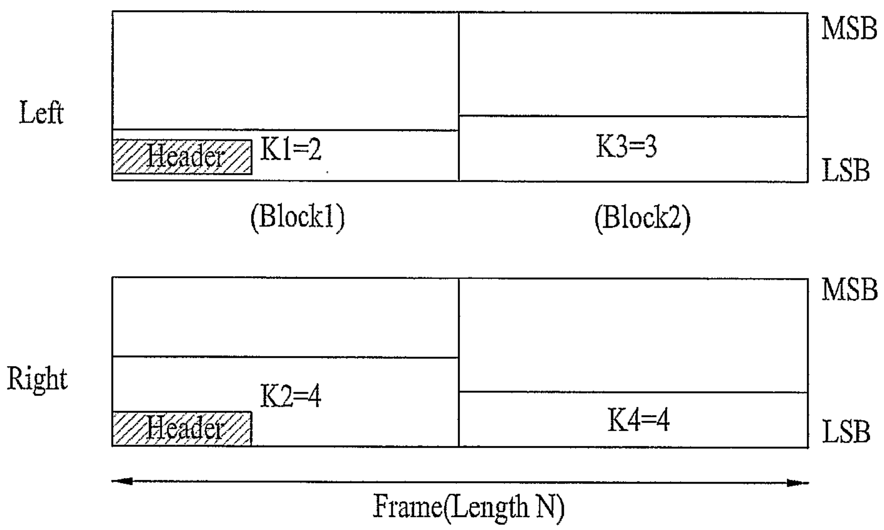

Left
Header
K1=2
K3=3
MSB
LSB
(Block1)
(Block2)
Right
K2=4
Header
K4=4
MSB
LSB
Frame(Length N)

FIG. 24

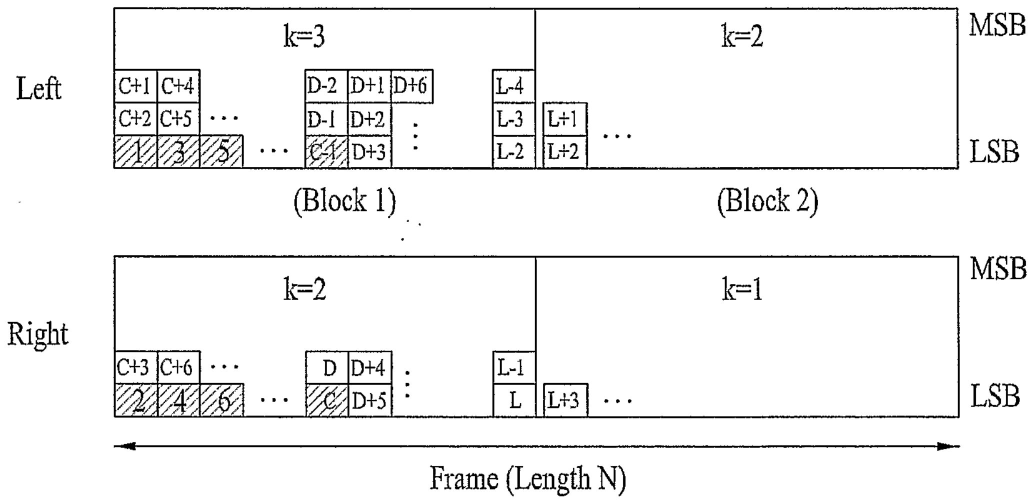

k=3
k=2
MSB
Left
C+1
C+4
D-2
D+1
D+6
L-4
C+2
C+5
D-1
D+2
L-3
L+1
1
3
5
C-1
D+3
L-2
L+2
LSB
(Block 1)
(Block 2)
k=2
k=1
MSB
Right
C+3
C+6
D
D+4
L-1
2
4
6
C
D+5
L
L+3
LSB
Frame (Length N)

FIG. 25

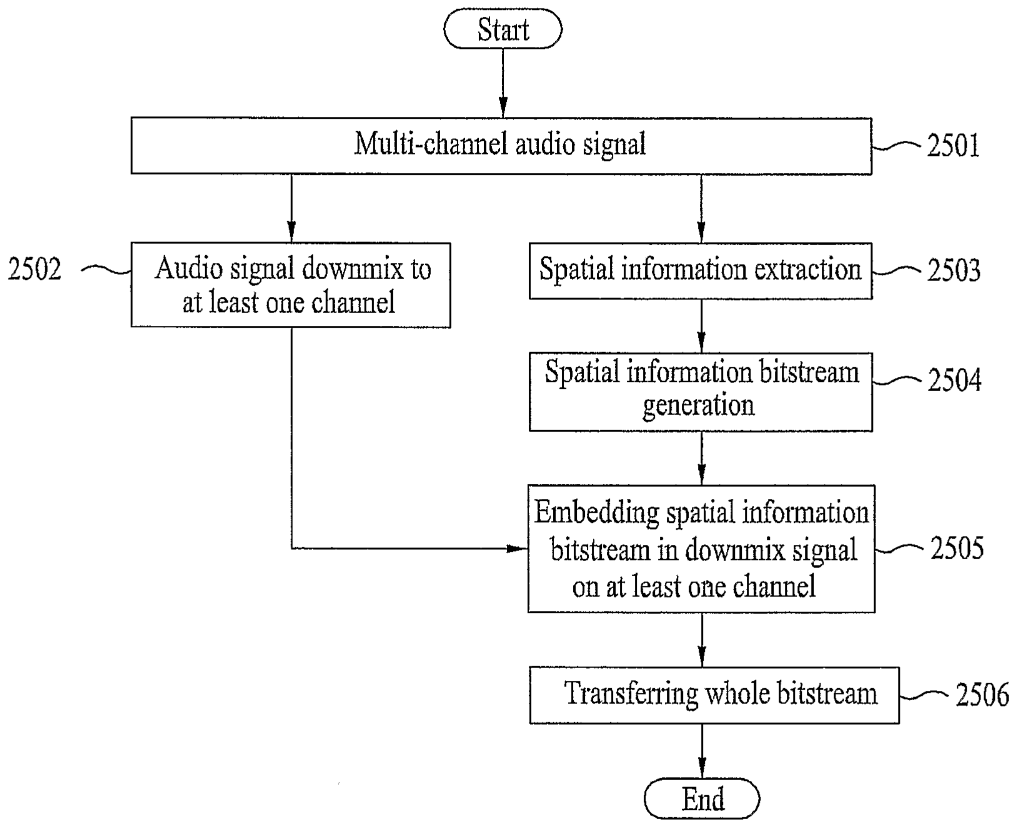
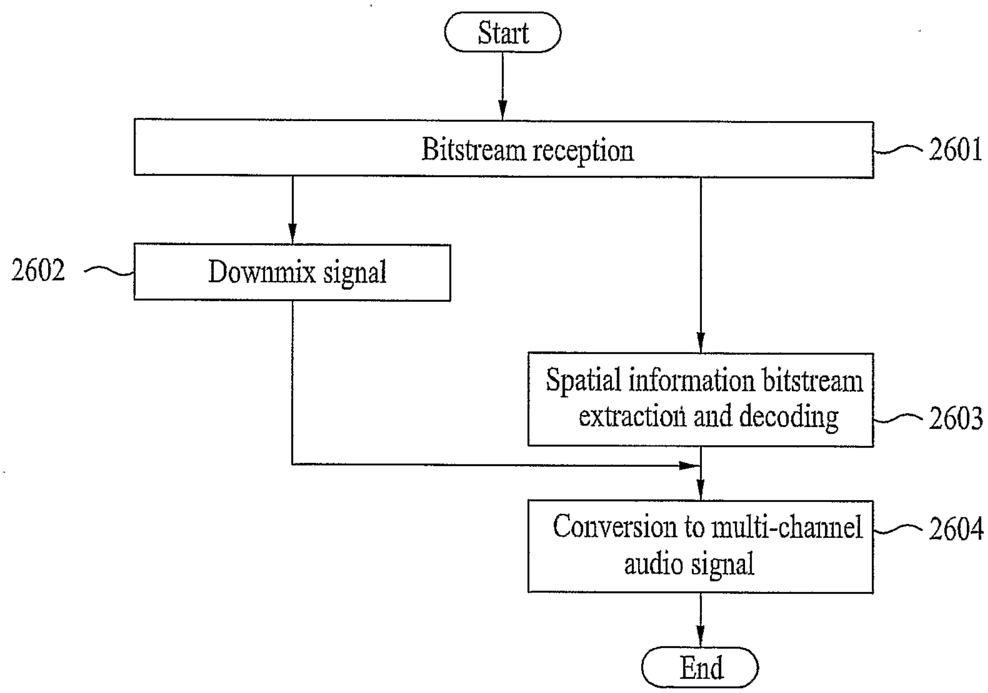

Start
Multi-channel audio signal
2501
2502
Audio signal downmix to at least one channel
Spatial information extraction
2503
Spatial information bitstream generation
2504
Embedding spatial information bitstream in downmix signal on at least one channel
2505
Transferring whole bitstream
2506
End

FIG. 26


Start
Bitstream reception
2601
2602
Downmix signal
Spatial information bitstream extraction and decoding
2603
Conversion to multi-channel audio signal
2604
End